

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Apr 20 16:14:23 2000; MasPar time 6236.81 Seconds
Tabular output not generated. 1390.567 Million cell updates/sec

Title: >US-09-249-003-1
Description: (1-2297) from US09249003.seq
Perfect Score: 2297
N.A. Sequence: 1 CGCGTGTGTTGCCAGTCCTT.....TAATCCATGTTTCCTTTCC 2297
Comp: GCCGACACACGGCTCAGAA.....ATTAGGTACAAAGAAAAGG

Scoring table:

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 4538634 seqs, 1887831982 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

1:em_est15 2:em_est23 3:em_est29 4:em_est3 5:em_est30
6:em_est31 7:em_est33 8:em_est34 9:em_est5 10:em_gss1
11:em_gss12

Database:

genbank-est15
12:gb_est1 13:gb_est10 14:gb_est11 15:gb_est12
16:gb_est13 17:gb_est14 18:gb_est15 19:gb_est16
20:gb_est17 21:gb_est18 22:gb_est19 23:gb_est2
24:gb_est20 25:gb_est21 26:gb_est22 27:gb_est23
28:gb_est24 29:gb_est25 30:gb_est26 31:gb_est27
32:gb_est28 33:gb_est29 34:gb_est3 35:gb_est30
36:gb_est31 37:gb_est32 38:gb_est33 39:gb_est34
40:gb_est35 41:gb_est36 42:gb_est37 43:gb_est38
44:gb_est39 45:gb_est4 46:gb_est40 47:gb_est41
48:gb_est42 49:gb_est43 50:gb_est44 51:gb_est5
52:gb_est6 53:gb_est7 54:gb_est8 55:gb_est9 56:gb_gss1
57:gb_gss10 58:gb_gss11 59:gb_gss12 60:gb_gss13
61:gb_gss14 62:gb_gss15 63:gb_gss2 64:gb_gss3 65:gb_gss4
66:gb_gss5 67:gb_gss6 68:gb_gss7 69:gb_gss8 70:gb_gss9

Statistics: Mean 12.223; Variance 2.355; scale 5.190

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	736	32.0	750	26	AT042325 oy37e10.x1 Soares_para	0.00e+00
2	721	31.4	752	27	A1089762 qa22e08.x1 NCI_CGAP.Br	0.00e+00
3	713	31.0	777	21	AA701678 z143h11.s1 Soares_feta	0.00e+00

ALIGNMENTS

RESULT	1	AI042325	750 bp	mrna	EST	30-JUN-1998
LOCUS	oy37e10.x1 Soares_parathyroid_tumor_NDHPA Homo sapiens cDNA clone					
DEFINITION	IMAGE:1668042 3' similar to gb:M58342 IDURONATE 2-SULFATASE					
	PRECUSOR (HUMAN);, mRNA sequence.					
ACCESSION	AI042325					
VERSION	AI042325.1	GI:3281519				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
	Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 750)					
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),					
	Tumor Gene Index					
JOURNAL	Unpublished (1997)					
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2151624.					
	Contact: Robert Strausberg, Ph.D.					
	Tel: (301) 496-1550					
	Email: Robert_Strausberg@nih.gov					
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima					
	Bonaldo, Ph.D.					
	cDNA Library Arrayed by: Greg Lennon, Ph.D.					
	DNA Sequencing by: Washington University Genome Sequencing Center					
	Clone distribution: NCI-CGAP clone distribution information can be					
	found through the I.M.A.G.E. Consortium/LLNL at:					

786 40	AI819115	wj79c02.x1	NCI_CGAP_Lu	0.00e+00
786 33	AI553770	tn28g04.x1	NCI_CGAP.Br	0.00e+00
694 37	AI744724	tn90d04.x1	NCI_CGAP.Co	0.00e+00
789 22	AA805711	ns42c05.s1	NCI_CGAP_GC	0.00e+00
653 50	AW173232	x185a05.x1	Soares_NFL	0.00e+00
704 19	AA553786	n135d03.s1	NCI_CGAP_Lu	0.00e+00
606 18	AA513794	nh89c05.r1	NCI_CGAP.Br	0.00e+00
644 13	AA131455	z132d04.r1	Soares_preg	0.00e+00
629 32	AI478887	tn25b02.x1	Soares_NFL	0.00e+00
576 14	AA236900	z843h10.s1	Soares_NHML	0.00e+00
566 38	AI769955	wj22f11.x1	NCI_CGAP_Ki	0.00e+00
551 31	AI418603	tg37f08.r1	Soares_NFL	0.00e+00
556 17	AA399142	z149f03.r1	Soares_ovar	0.00e+00
604 26	AI033665	ov22g07.x1	Soares_para	0.00e+00
651 20	AA642413	ns27g11.s1	NCI_CGAP_GC	0.00e+00
552 18	AA506162	n114a10.s1	NCI_CGAP.Co	0.00e+00
552 32	AI471587	ta15f06.x1	NCI_CGAP_Ly	0.00e+00
496 27	AI143191	qb72a12.x1	Soares_feta	0.00e+00
712 39	AI798738	we92a07.x1	Soares_NFL	0.00e+00
986 32	AI524108	th09c01.x1	NCI_CGAP_CL	0.00e+00
526 48	AW160323	au66a10.y1	Schneider_f	0.00e+00
570 29	AI271813	qj88a08.x1	NCI_CGAP_Ki	0.00e+00
589 31	AI400170	tg67h06.x1	Soares_NHML	0.00e+00
525 53	W90003	zh69b11.s1	Soares_feta	0.00e+00
504 25	AA935580	on13d04.s1	NCI_CGAP_Lu	0.00e+00
615 34	H06357	y179h02.r1	Soares_infa	0.00e+00
469 30	AI359729	qy28f03.x1	NCI_CGAP.Br	0.00e+00
511 28	AI186134	qe34h05.s1	Soares_feta	0.00e+00
516 29	AI283859	qk46c03.x1	NCI_CGAP.Co	0.00e+00
466 28	AI174207	on01b09.x1	Stratagene	0.00e+00
480 29	AI292239	qm76a02.x1	Soares_plac	0.00e+00
453 27	AI096538	qb61d01.x1	NCI_CGAP.Br	0.00e+00
443 30	AI359745	qy28h03.x1	NCI_CGAP.Br	0.00e+00
455 44	AW134717	UI-H-Bil-abq-a-11-0-UI		0.00e+00
437 30	AI359737	qy28g03.x1	NCI_CGAP.Br	0.00e+00
519 21	AA700596	z141b07.s1	Soares_feta	0.00e+00
479 51	N26477	yy61b05.s1	Soares_mela	0.00e+00
452 25	AA993642	ot84g10.s1	Soares_tota	0.00e+00
460 13	AA131408	z132d04.s1	Soares_preg	0.00e+00
467 36	AI678627	tu84h05.x1	NCI_CGAP_Ga	0.00e+00
517 34	H08043	y191g05.r1	Soares_infa	0.00e+00

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 474.

FEATURES

source

BASE COUNT	164 a	231 c	177 g	176 t	2 others
ORIGIN					
Query Match	32.0%;	Score 736;	DB 26;	Length 750;	
Best Local Similarity	99.2%;	Pred. No. 0.00e+00;			
Matches	745;	Conservative	0;	Mismatches	5;
				Indels	1;
				Gaps	1;
Db	1	CGCGCCGCTCGAAGCCGAATGCGCCACCCCGGACCGCGGAGCGCTTCTCTGGCTGG	60		
QY	105	CGCGCCGCTCGAAGCCGAATGCGCCACCCCGGACCGCGGAGCGCTTCTCTGGCTGG	164		
Db	61	GTCTGGTCTGAGCTCGCTCGCTCGCTCGGATCCGAAACGAGGCCAACTCGACCA	120		
QY	165	GTCTGGTCTGAGCTCGCTCGCTCGGATCCGAAACGAGGCCAACTCGACCA	224		
Db	121	CAGATGCTCTGAACGTTCTTCTCATCATCTGTTGATGACCTGCGCCCTCCCTGGCGTGT	180		
QY	225	CAGATGCTCTGAACGTTCTTCTCATCATCTGTTGATGACCTGCGCCCTCCCTGGCGTGT	284		
Db	181	ATGGGGATAAGCTGGTGAGGTCCCCAAATATGACCAACTGGCATCCACAGCCCTCCTCT	240		
QY	285	ATGGGGATAAGCTGGTGAGGTCCCCAAATATGACCAACTGGCATCCACAGCCCTCCTCT	344		
Db	241	TCCAGATGCTTTGCGCAGCAAGCAGTGTGCGCCCGAGCGGGTTCCTTCCTCACTG	300		
QY	345	TCCAGATGCTTTGCGCAGCAAGCAGTGTGCGCCCGAGCGGGTTCCTTCCTCACTG	404		
Db	301	GCAGGAGACCTGCACACACCGCCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTG	360		
QY	405	GCAGGAGACCTGCACACACCGCCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTG	464		
Db	361	GAAACTTCTCACCATCCCCAGPACTTCAAGGAGAAATGGCTATGTGACCATGTCGGTGG	420		
QY	465	GAAACTTCTCACCATCCCCAGPACTTCAAGGAGAAATGGCTATGTGACCATGTCGGTGG	524		
Db	421	GAAAGTCTTTCACCCCTGGGATATCTTCTAACCATAGTATGATGATCTCCGATPAGCTGGT	480		
QY	525	GAAAGTCTTTCACCCCTGGGATATCTTCTAACCATAGTATGATGATCTCCGATPAGCTGGT	584		
Db	481	CTTTTCCACCTTATCATCCTTCCCTCTGAGAGAGTATGAAACACTAAGACATGTCGAGGC	540		
QY	585	CTTTTCCACCTTATCATCCTTCCCTCTGAGAGAGTATGAAACACTAAGACATGTCGAGGC	644		
QY	541	CAGATGAGAACTCCATGCCAACCTGCTNTGCCCTGTGGATGTGCTGGATGTTCCCGAGG	600		

QY	645	CAGATGGAGAACTCCATGCCAACCTGCTTTGGCCCTGTGGGATGTGCTGGATGTTCCCGAGG	704
Db	601	GCACCTTTGCCTGCAAAACAGAGACACTGAGCAAGCCATACAGTTGTTGGAAAGATGAAA	660
QY	705	GCACCTTTGCCTGCAAAACAGAGACACTGAGCAAGCCATACAGTTGTTGGAAAGATGAAA	764
Db	661	CGTCAGCCAGTCCCTTTCTTCTCTG-CGCTTGGGATCATAAAGCCACACATNCCCTTCAGAT	719
QY	765	CGTCAGCCAGTCCCTTTCTTCTCTGCGCTTGGGATCATAAAGCCACACATCCCCCTTCAGAT	824
Db	720	ACCCCCAAGAATTTCAAGAAGTTGTATCCCTT	750
QY	825	ACCCAAAGGAATTTCAAGAAGTTGTATCCCTT	855
RESULT	2		
LOCUS	AI089762	752 bp	mRNA
DEFINITION	ga22e08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1687526 3' similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA sequence.	EST	01-OCT-1998
ACCESSION	AI089762		
VERSION	AI089762.1	GI:3428821	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	NCI (bases 1 to 752)		
TITLE	NCI/NINDS-CGAP project://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index		
JOURNAL	Unpublished (1998)		
COMMENT	On Oct 20, 1997 this sequence version replaced gi:2519916. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbr/image/image.html		

Insert Length: 1268 Std Error: 0.00
Seq primer: ~40m13 fwd. ET from Amersham
High quality sequence stop: 425.

FEATURES
source

source

```

i.: 152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1687526"
/clone_lib="NCL CGAP_Brn23"
/tissue_type="Glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTGTACCATCTGAACTGGAGCGCCGATATCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Ronaldo."

```

BASE COUNT		SOARES AND M.FALTIMA BONDAGE.		1 others	
167 a	230 c	177 q	177 t	177 t	1 others

ORIGIN

Query Match 31.4%; Score 721; DB 27; Length 752;
Best Local Similarity 98.5%; Pred.No. 0.00e+00;
Matches 741; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

b 481 TCACGGGATTAC CAGGGAGGTACGGATCCTCTTCCAAGTCACGGAATCGAAATGCTTC 540

ACCESSION AA701678
 VERSION AA701678.1 GI:2704843
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 777)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenger, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
 TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)

Cp 1500 TCACGGGATTACACGGAGGTACGGATCCTCTTCCAAAGTCACGGAATCGAAATGCTTC 1441

Db 541 AGAAGGTTCTTGCCTTCTGTCACAGCTCAACGTGAATGAAGAACGGGCGCAGGTT 600

Cp 1440 AGAAGGTTCTTGCCTTCTGTCACAGCTCAACGTGAATGAAGAACGGGCGCAGGTT 1381

Db 601 GGAACCTGCAGTCTGCAAGTTACAGCGCTTGGGAAAGAGACAAAGTCCCAACA 660

Cp 1380 GGAACCTGCAGTCTGCAAGTTACAGCGCGT - GGGAAAGAGACAAAGTTCACAA -A 1323

Db 661 GGTTCATGGATTGCGCTGCTGCCATTAACCTTGAGGCGGAATCAAAAGGTCGAGG 720

Cp 1322 GTCCATGGATTGCGCTGCTGCCATTC - ATCAACTGTGAGGCGGAATCAAAAGGTCGAGG 1264

Db 721 TAAGGGAAAGCTTCTCGCTGCTGTCGCGAAGTGAAGACGTTCTTCCAAAGAACATA 777

Cp 1263 TAAGGGAAAGCTTCTCGCTGCTGTCGCGAAGTGAAGCGCTTCTTCCA - GGAACATA 1208

RESULT 4
 LOCUS wj79c02.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2409026 3',
 DEFINITION similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
 sequence.
 ACCESSION AF1819115
 VERSION AF1819115.1 GI:5438118
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 786)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189120.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/dbp/image/image.html

Seq primer: -400P from Gibco
 High quality sequence stop: 434.
 FEATURES
 source Location/Qualifiers
 1..786
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2409026"
 /clone_lib="NCI_CGAP_Lu19"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 175 a 242 c 182 g 179 t 8 others

BASE COUNT
 ORIGIN

Query Match 30.8%; Score 708; DB 40; Length 786;
 Best Local Similarity 96.8%; Pred. No. 0.00e+00;
 Matches 766; Conservative 0; Mismatches 17; Indels 8; Gaps 8;

Db 1 GCTGGTCTGTGTTCTGAGCTCGTCTGCGTCCGCTCGATCCGAAACCGACGCCAATC 60

Qy 160 GCTGGTCTGTGTTCTGAGCTCGTCTGCGTCCGCTCGATCCGAAACCGACGCCAATC 219

Db 61 GACCACAGATGCTCTGAAGTTCTTCTCATCATCTGTTGATGACCTGCGCCCTCCCTGGG 120

Qy 220 GACCACAGATGCTCTGAAGTTCTTCTCATCATCTGTTGATGACCTGCGCCCTCCCTGGG 279

Db 121 CTGTTATGGGGATAAGCTGGTGGTCCCAAAATATTGACCAACTGGCATCCACAGCCT 180

Qy 280 CTGTTATGGGGATAAGCTGGTGGTCCCAAAATATTGACCAACTGGCATCCACAGCCT 339

Db 181 CCTCTTCCAGATGCTTTGGCGCAGCAGTGTGCGCCCGCGCGGCTTCTTTCTCT 240

Qy 340 CCTCTTCCAGATGCTTTGGCGCAGCAGTGTGCGCCCGCGCGGCTTCTTTCTCT 399

Db 241 CACTGGCAGGAGACCTGACACACCGCGCTGTACGACTTCAACTCTCTACTGGAGGTGCA 300

Qy 400 CACTGGCAGGAGACCTGACACACCGCGCTGTACGACTTCAACTCTCTACTGGAGGTGCA 459

Db 301 CGCTGGAACCTTCTCCACCATCCCGCAGTACTCAAGGAGAATGGCTATGTGACCATGTC 360

Qy 460 CGCTGGAACCTTCTCCACCATCCCGCAGTACTCAAGGAGAATGGCTATGTGACCATGTC 519

Db 361 GGTGGGAAAAGTCTTTCACCTGGGATATCTTCTTAACCATACCGATGATTCCTCGATAG 420

Qy 520 GGTGGGAAAAGTCTTTCACCTGGGATATCTTCTTAACCATACCGATGATTCCTCGATAG 579

Db 421 CTGCTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACATAAGACATGTCG 480

Qy 580 CTGCTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACATAAGACATGTCG 639

Db 481 AGGGCCAGATGGAGAATCCATGCCCAACCTGCTNTGCCCTGGTGGATGTGCTG - ATGTTC 539

Qy 640 AGGGCCAGATGGAGAATCCATGCCCAACCTGCTNTGCCCTGGTGGATGTGCTG - ATGTTC 698

Db 540 CCGAGGGCACTTGGCTGACAAACAGAGACACTGAGNCAGCCATACAGTTGTTGGAAAGA 599

Qy 699 CCGAGGGCACTTGGCTGACAAACAGAGACACTGAGNCAGCCATACAGTTGTTGGAAAGA 758

Db 600 TGAAAACGTACCCAGTCTTCTTCNCTGCGCGTGGGTATCATAGCCACACATNCCCT 659

Qy 759 TGAAAACGTACCCAGTCTTCTTCNCTGCGCGTGGGTATCATAGCCACACATNCCCT 818

Db 660 TCAGATACCCCAAGAAATTTCCAGAAGT - GTATCCCTGGAGAACATCACCNTGGCCCC 718

Qy 819 TCAGATACCCCAAGAAATTTCCAGAAGT - GTATCCCTGGAGAACATCACC - CTGCCCC 876

Db 719 CGATCCCGAG - TNCCTGTGGGTACCCCTGTG - CCTACAAACCCCTGGATG - ACATCAN 775

Qy 877 CGATCCCGAGGTCCCTGTGGGTACCCCTGTGGGTACAAACCCCTGTGGATGGACATCAG 936

Db 776 GCAACGGNAAG 786

Qy 937 GCAACGGNAAG 947

RESULT 5
 LOCUS AI553770 786 bp mRNA EST 12-MAY-1999
 DEFINITION tn28g04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2168982 3',
 similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
 sequence.
 ACCESSION AI553770
 VERSION AI553770.1 GI:4486133
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 786)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1997)
On Mar 20, 1998 this sequence version replaced gi:2980487.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1027 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 466
POLYA=No.

FEATURES

Location/Qualifiers
1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2168982"
/clone_lib="NCI CGAP Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(GT) primer [5'
TGTACCAATCTGAGTGGAGCGCGCATGAGTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 173 a 242 C 188 G 182 T 1 others
ORIGIN

Query Match 30.4%; Score 699; DB 33; Length 786;
Best Local Similarity 97.1%; Pred. No. 0.00e+00;
Matches 761; Conservative 0; Mismatches 15; Indels 8; Gaps 8;

Db 1 GCGGCGGTCGAAGCGGAAATGCGGCCACCGCGACTTTCGAGGCGCTTCTCTGCTGG 60
Qy 105 GCGGCGGTCGAAGCGGAAATGCGGCCACCGCGAGCGGCGCTTCTCTGCTGG 164
Db 61 GCTCGTTCTGAGCTCCGTCCTGGTGGCCCTGGATCCGAAGCGGCAACTCGACCA 120
Qy 165 GCTCGTTCTGAGCTCCGTCCTGGTGGCCCTGGATCCGAAGCGGCAACTCGACCA 224
Db 121 CAGATGCTCTGAACCTTCTTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTT 180
Qy 225 CAGATGCTCTGAACCTTCTTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTT 284
Db 181 ATGGGGATAAGCTGGTGGAGTCCCAAAATATTGACCACTGGCATCCCAAGCCTCTCT 240
Qy 285 ATGGGGATAAGCTGGTGGAGTCCCAAAATATTGACCACTGGCATCCCAAGCCTCTCT 344
Db 241 TCCGAATGCCCTTTCGGCAGCAGAGTGGCGCCCGGAGCGCGTTCCTTCCTCACTG 300
Qy 345 TCCGAATGCCCTTTCGGCAGCAGAGTGGCGCCCGGAGCGCGTTCCTTCCTCACTG 404
Db 301 CGAGGAGACCTGACACCGCCCTGTAGCACTTCAACTCCTACTGGAGGGTGCAGCTG 360
Qy 405 CGAGGAGACCTGACACCGCCCTGTAGCACTTCAACTCCTACTGGAGGGTGCAGCTG 464

Db 361 GAAACTTCTCCACCATCCCCAGTACTTCAAGAGAGATGGCTATGTGACCATGTGCGTGG 420
Qy 465 GAAACTTCTCCACCATCCCCAGTACTTCAAGAGAGATGGCTATGTGACCATGTGCGTGG 524
Db 421 GAAAGTCTTTCACCCCTGGGATATCTTCTAACCATACCGATGATTCCTCCGTATAGCTGGT 480
Qy 525 GAAAGTCTTTCACCCCTGGGATATCTTCTAACCATACCGATGATTCCTCCGTATAGCTGGT 584
Db 481 CTTTTCACCTTATCATCTCTTCTCTGAGAGTATGAAAACACATAGACATGTCCGAGGGC 540
Qy 585 CTTTTCACCTTATCATCTCTTCTCTGAGAGTATGAAAACACATAGACATGTCCGAGGGC 644
Db 541 CAGATGAGAACTCCATGCCAACCTGCTTTCCTCTGAGATGTCTGGGATGTTCGAGG 600
Qy 645 CAGATGAGAACTCCATGCCAACCTGCTTTCCTCTGAGATGTCTGGGATGTTCGAGG 704
Db 601 GCACCCCTGCTGACAA-CAGAGCAGCTGAGCAAGCCATACAGTTGTGGAAAAAGATGAAA 659
Qy 705 GCACCT-TGCCTGACAAACAGAGCAGCTGAGCAAGCCATACAGTTGTGGAAAAAGATGAAA 763
Db 660 ACGTCAAGCAGTCTTCTTCTTCTGCGCTTGGGATGTCTGGGATGTTCGAGG 717
Qy 764 ACGTCAAGCAGTCTTCTTCTTCTGCGCTTGGGATGTCTGGGATGTTCGAGG 823
Db 718 TACCCCAAGGAAT-CAGAGTGTG-ATCCTTTG-AGAACAT-ACCCTGGCCCCCGATCCC 773
Qy 824 TACCCCAAGGAATTCAGAGTGTGTATCTCTTGGAGAACATCACCTTGGCCCCCGATCCC 883
Db 774 AAG 777
Qy 884 GAG 887

RESULT 6
LOCUS AI744724 694 bp mRNA EST 21-JUN-1999
DEFINITION wf90404.x1 NCI-CGAP Co3 Homo sapiens cDNA clone IMAGE:2362855 3'
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
sequence.
ACCESSION AI744724
VERSION AI744724.1 GI:5112934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135683.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 462.
Location/Qualifiers
1..694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2362855"
/clone_lib="NCI-CGAP Co3"
/sex="pooled"
/tissue_type="colon"

FEATURES
source

/lab_host="DH10B"									
/note=vector: pf7f3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pf7f3 vector. Library went through one round of normalization.									
BASE COUNT	150 a	215 c	164 g	163 t	2 others				
ORIGIN									
Query Match	29.6%	Score 680;	DB 37;	Length 694;					
Best Local Similarity	99.4%	Pred. No. 0.00e+00;							
Matches 692;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;					
Db	1	AAATGCGCGCCACCGCGGACCGCGGAGCGCTTCTGTGGCTGGTCTGGTCTGAGCTCCG	60						
Qy	123	AAATGCGCGCCACCGCGGACCGCGGAGCGCTTCTGTGGCTGGTCTGGTCTGAGCTCCG	182						
Db	61	TCTGCGTGGCCCTCGGATTCGGAATCGAAGCGAGGCCAACTCGACCACAGATGCTCTGAACGTTTC	120						
Qy	183	TCTGCGTGGCCCTCGGATTCGGAATCGAAGCGAGGCCAACTCGACCACAGATGCTCTGAACGTTTC	242						
Db	121	TTCTCATCATCGTGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGGATAAAGCTGGTGA	180						
Qy	243	TTCTCATCATCGTGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGGATAAAGCTGGTGA	302						
Db	181	GGTCCCCAAATATTGACCACTGGCATGCCACAGCGCTCTCTCCAGAATGCCCTTTGGCG	240						
Qy	303	GGTCCCCAAATATTGACCACTGGCATGCCACAGCGCTCTCTCCAGAATGCCCTTTGGCG	362						
Db	241	AGCAAGCAGTGTGGCCCCCGAGCGCGTTCCTTCCCTCACTGGCAGGAGACCTGACACCA	300						
Qy	363	AGCAAGCAGTGTGGCCCCCGAGCGCGTTCCTTCCCTCACTGGCAGGAGACCTGACACCA	422						
Db	301	CCCGCCTGTAGCACTTCAACTCCTACTGGAGGTGTCAGCTGGAACCTTCCACCATCC	360						
Qy	423	CCCGCCTGTAGCACTTCAACTCCTACTGGAGGTGTCAGCTGGAACCTTCCACCATCC	482						
Db	361	CCCAAGTACTTCAAGGAGAATGGCTATGTGACCATGTGGTGGGAAAAGTCTTTCACCCCTG	420						
Qy	483	CCCAAGTACTTCAAGGAGAATGGCTATGTGACCATGTGGTGGGAAAAGTCTTTCACCCCTG	542						
Db	421	GGATATCTTCTTAACCATACCGATGATTCCTCGTATAGCTGGTCTTTTCCACCTTATCATC	480						
Qy	543	GGATATCTTCTTAACCATACCGATGATTCCTCGTATAGCTGGTCTTTTCCACCTTATCATC	602						
Db	481	CTTCCTCTGAGAAGTATGAAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCATG	540						
Qy	603	CTTCCTCTGAGAAGTATGAAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCATG	662						
Db	541	CCAACCTGCTNTGCCCTGTGGATGCTGTGGATGTTCCCGAGGGCACCTTGCCTGCACAAAC	600						
Qy	663	CCAACCTGCTNTGCCCTGTGGATGCTGTGGATGTTCCCGAGGGCACCTTGCCTGCACAAAC	722						
Db	601	AGAGCACTGAGCAAGCCATACAGTGTGTGGAAAAGATGAAAACGTCAGCCAGTCCCTTCT	660						
Qy	723	AGAGCACTGAGCAAGCCATACAGTGTGTGGAAAAGATGAAAACGTCAGCCAGTCCCTTCT	782						
Db	661	TCCTGGCGGTTGG--TATCATTA-GCCACACATNCCCT	694						
Qy	783	TCCTGGCGGTTGGTATCATTAAGCCACACATCCGCT	818						

RESULT	7				
LOCUS	AA805711	789 bp	mRNA	EST	18-FEB-1998
DEFINITION	ns42c05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1186280 3' similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN); mRNA sequence.				
ACCESSION	AA805711				
VERSION	AA805711.1	GI:2874461			
KEYWORDS	EST.				

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 789)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gi:1797466.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1068 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 490.
Location/Qualifiers
1. .789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1186280"
/clone_lib="NCI-CCGAP-GCBI"
/issue_type="germinal center B cell"
/lab_host="DH10b"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCACTGAAGTGGAGCGCCGCTCATTTTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 253 a 140 c 148 g 248 t
ORIGIN

Query Match 29.6%; Score 681; DB 22; Length 789;
Best Local Similarity 96.8%; Pred. No. 0.00e+00;
Matches 751; Conservative 0; Mismatches 16; Indels 9; Gaps 9;

Db 22 GAAAGAACAATGGATTACATGTTTACATAAATTTTGAATATATAGGGCTCATAGAT 81
|||||
Cp 2296 GAAAGAACAATGGATTACATGTTTACATAAATTTTGAATATATAGGGCTCATAGAT 2237
|||||
Db 82 ATATTATTAATTATAACAAAATAATCAGGAACCTAGTGTGGTGCCTAGTTGATATAT 141
|||||
Cp 2236 ATATTATTAATTATAACAAAATAATCAGGAACCTAGTGTGGTGCCTAGTTGATATAT 2177
|||||
Db 142 GATTACTTTTGAATGCACCTAAATCCCAAAATATGAAGTATTCCTTTGCTGATATATG 201
|||||
Cp 2176 GATTACTTTTGAATGCACCTAAATCCCAAAATATGAAGTATTCCTTTGCTGATATATG 2117
|||||
Db 202 TTATGTTTGGTTATTTATGTTATGCTTCGATCCAAAGGTATGACATAACTTCAGTTTGT 261
|||||
Cp 2116 TTATGTTTGGTTATTTATGTTATGCTTCGATCCAAAGGTATGACATAACTTCAGTTTGT 2057
|||||
Db 262 TTGCTTTGTTATTTATTCAGTAAATAAGCCGTAAGTCTGTTTTAAAGAGGGAATTAATA 321
|||||
Cp 2056 TTGCTTTGTTATTTATTCAGTAAATAAGCCGTAAGTCTGTTTTAAAGAGGGAATTAATA 1997
|||||

Db 322 AAAAAGTGGTCAATTACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCT 381
|||||
Cp 1996 AAAAAGTGGTCAATTACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCT 1937
|||||
Db 382 GTTACATATTTCTAGCGCAAAATGTTGATGATGATGTTGATTAATAGCCCTCAGGCTGC 441
|||||
Cp 1936 GTTACATATTTCTAGCGCAAAATGTTGATGATGATGTTGATTAATAGCCCTCAGGCTGC 1877
|||||
Db 442 TTCCAATATTATGGTATATCAAAACACAGCTCTAATCTCTCTCACCAGCTGA 501
|||||
Cp 1876 TTCCAATATTATGGTATATCAAAACACAGCTCTAATCTCTCTCACCAGCTGA 1817
|||||
Db 502 AGGAGCACATCACATTTGCCATCATCGTTGGCAAACTCAAGGCATCAACAACCTGGAA 561
|||||
Cp 1816 AGGAGCACATCACATTTGCCATCATCGTTGGCAAACTCAAGGCATCAACAACCTGGAA 1757
|||||
Db 562 AAGATCTCCCTCTGGGAATCATATATATATGTGATCCTGCAATGGGTCAAGATCCCC 621
|||||
Cp 1756 AAGATCTCCCTCTGGGAATCATATATATATGTGATCCTGCAATGGGTCAAGATCCAC 1697
|||||
Db 622 AAA-TAGGTTCCCTGACATGGAATTTCCGAAAGTAGCTGGAATCC-TCCGGATTGA 679
|||||
Cp 1696 AAAATACAGTTCCCTGCA-TGGATGTGAGAAAGTTAGCTAGAAATTCATCAGGATTGA 1638
|||||
Db 680 AGCCA-CCACAC-GTATACC-ATGGTCCATGCTGCGTATGGAATAGCC-ATGATCTTTA 735
|||||
Cp 1637 AGCCAACCCACACAGTATACCTATAGTGTGGTGGTATGGAATAGCCCATGATCTTTA 1578
|||||
Db 736 TATCTTT-AAACTCGCTGGTGCAGAAATTCCTCA-TGAGGATGCTGAGGCGCGGGGA 789
|||||
Cp 1577 TATCTTTAAACTCGCTGGTGCAGAAATTCCTCACTGAGGATGCTGAGGCGCGGGGA 1522
|||||

RESULT 8
LOCUS AW173232 653 bp mRNA EST 16-NOV-1999
DEFINITION x185a05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663984 3' similar to gb:M58342 IDURONATE 2-SULFATASE
PRECURSOR (HUMAN); mRNA sequence.
AW173232
ACCESSION AW173232.1 GI:6439180
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 653)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
On Apr 30, 1999 this sequence version replaced gi:5866661.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -40up from Gibco
High quality sequence stop: 440.
Location/Qualifiers
1. .653
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663984"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 143 a 196 c 154 g 157 t 3 others

Query Match 27.8%; Score 638; DB 50; Length 653;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;

Matches 647; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 1 GCTGGGCTGTTCTGAGCTCCGCTCGCTCGCCCTCGGATCGAAACGAGGCGCAACTC 60
|||||

Qy 160 GCTGGGCTGTTCTGAGCTCCGCTCGCTCGCCCTCGGATCGAAACGAGGCGCAACTC 219
|||||

Db 61 GACCACAGATGCTCTGAACGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 120
|||||

Qy 220 GACCACAGATGCTCTGAACGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 279
|||||

Db 121 CTGTTATGGGGATAAGCTGCTGAGGTCCTCCCAATATTTGACCAACTGCCATCCACAGCCT 180
|||||

Qy 280 CTGTTATGGGGATAAGCTGCTGAGGTCCTCCCAATATTTGACCAACTGCCATCCACAGCCT 339
|||||

Db 181 CTTCTTCCAGAATGCTTTGCGCAGCAAGCAGTGTGCGCCCGGAGCGCGCTTTCTTCTCT 240
|||||

Qy 340 CTTCTTCCAGAATGCTTTGCGCAGCAAGCAGTGTGCGCCCGGAGCGCGCTTTCTTCTCT 399
|||||

Db 241 CACTGCGAGGAGACCTGACACACCGCCCTGTACGACTTCAACTCTCTTCTGAGGGTGCA 300
|||||

Qy 400 CACTGCGAGGAGACCTGACACACCGCCCTGTACGACTTCAACTCTCTTCTGAGGGTGCA 459
|||||

Db 301 CGCTGGAATCTCTCCACCATCCCGCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTC 360
|||||

Qy 460 CGCTGGAATCTCTCCACCATCCCGCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTC 519
|||||

Db 361 GGTGGGAAAAGTCTTTCACCCCTGGGATATCTTCAACCATACCGATGATTTCTCGGTATAG 420
|||||

Qy 520 GGTGGGAAAAGTCTTTCACCCCTGGGATATCTTCAACCATACCGATGATTTCTCGGTATAG 579
|||||

Db 421 CTGGTCTTTTCCACCTTATCATCTCTCTGAGAGTATGAAACACTAAGACATGTCG 480
|||||

Qy 580 CTGGTCTTTTCCACCTTATCATCTCTCTGAGAGTATGAAACACTAAGACATGTCG 639
|||||

Db 481 AGGGCCAGATGGAGAACTCCATGCCAACCTGCTNTGCCCTGTGGATGTGTTGATGTTCC 540
|||||

Qy 640 AGGGCCAGATGGAGAACTCCATGCCAACCTGCTNTGCCCTGTGGATGTGTTGATGTTCC 699
|||||

Db 541 CGAGGCGACCTTGCCTGACAAACAGAGCACTGAGCAAGCCATACANGTGTGGAAAAGAT 600
|||||

Qy 700 CGAGGCGACCTTGCCTGACAAACAGAGCACTGAGCAAGCCATACANGTGTGGAAAAGAT 759
|||||

Db 601 GATAAGCTCAGCAGTCTTCTCTCTGCGCGCTGG-TATCANTAGCCACAT 653
|||||

Qy 760 GAAAACGTGAGCAGTCTTCTCTCTGCGCGCTGG-TATCANTAGCCACAT 813
|||||

RESULT 9

LOCUS AA553786 704 bp mRNA EST 12-AUG-1997

DEFINITION ni35d03.s1 NCI-CGAP_Lu1 Homo sapiens cDNA clone IMAGE:978821 3'

similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN); mRNA

sequence.

ACCESSION AA553786

VERSION AA553786.1 GI:2324325

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 704)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
On Apr 14, 1993 this sequence version replaced gi:693458.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 392.

FEATURES
source

1..704
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:978821"
/clone_lib="NCI-CGAP_Lu1"
/tissue_type="lung tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: Bluescript SK-; Site_1: EcoRI;
Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
Bulk lung tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3'
3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
Average insert size: 1.1 kb."
211 a 163 c 161 g 169 t

BASE COUNT
ORIGIN

Query Match 27.8%; Score 639; DB 19; Length 704;
Best Local Similarity 96.9%; Pred. No. 0.00e+00;
Matches 686; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
Db 1 ACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTGGCTGTTCACATATTCTCAGG 60
|||||
Cp 1980 ACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTGGCTGTTCACATATTCTCAGG 1921
|||||
Db 61 CCAATTGTTGATCATATTGTTGGATTAACTAGCCCTCAGGCTGCTCCAAATTATTGGGT 120
|||||
Cp 1920 CCAATTGTTGATCATATTGTTGGATTAACTAGCCCTCAGGCTGCTCCAAATTATTGGGT 1861
|||||
Db 121 AATCAGAAAGCAGCAGCTCTAATCTCTCTCACCAGCTGGAGGAGCACATCATAT 180
|||||
Cp 1860 AATCAGAAAGCAGCAGCTCTAATCTCTCTCACCAGCTGGAGGAGCACATCATAT 1801
|||||
Db 181 TTGCCATCCATGGTTGGCAAAACTCAAGGCATCAACAACTGGAAGATCTCCACCTTGG 240
|||||
Cp 1800 TTGCCATCCATGGTTGGCAAAACTCAAGGCATCAACAACTGGAAGATCTCCACCTTGG 1741
|||||
Db 241 GAATCATATATACATATTGTGATCCAGCAATGGGTGAGATCCCAAAATACAGTTCCTCCT 300
|||||
Cp 1740 GAATCATATATACATATTGTGATCCAGCAATGGGTGAGATCCCAAAATACAGTTCCTCCT 1681
|||||
Db 301 GCATGGATGCAGAAAGTTAGCTAGAAAATTCATCAGATTGAGCCCAACCCACAGTA 360
|||||
Cp 1680 GCATGGATGCAGAAAGTTAGCTAGAAAATTCATCAGATTGAGCCCAACCCACAGTA 1621
|||||
Db 361 TACCTATAGTCTATGGTGGCGTATGGAATAGCCATGATCTTTATATCTTTTAAACTCGGC 420
|||||
Cp 1620 TACCTATAGTCTATGGTGGCGTATGGAATAGCCATGATCTTTATATCTTTTAAACTCGGC 1561
|||||
Db 421 TTGTTCAGAAATTCACCTGAGGATGCTCTGAAGCCGGGATAGTACTGGCTATAGGCAATCAGT 480
|||||
Cp 1560 TTGTTCAGAAATTCACCTGAGGATGCTCTGAAGCCGGGATAGTACTGGCTATAGGCAATCAGT 1501
|||||
Db 481 TCACGGGGAATTCACGGGAGGTACCGATCCTACTTCCAGTCACGGATCGAAATGCTT 540
|||||
Cp 1500 TCACGGGGA-TTACCGGGAGGTACCGATCCTCTTCCAAAGTCACGGATCGAAATGCTT 1442
|||||

Db 541 CAGAGGTTCTTGGCTTCTCTGCAAGGCTCAACGTGAATGAAGAACGGGCGAGG 600
|||||
Cp 1441 CAGAGGTTCTTGGCTTCTCTGCAAGGCTCAACGTGAATGAAGAACGGGCGAGG 1382
|||||
Db 601 TGGACCC-GCAGTCTGCA-GTCCAGCCAGCTGGGAGAAAGACACAGTTTCCCA-G 657
|||||
Cp 1381 TGGAACTGCACTGCTGCAAGTCCAGCCAGCTGGGAGAAAGACACAGTTTCCCAAG 1322
|||||
Db 658 GTGCAGGAGTTCCGCTGCTGCTCCATCAATGGTGAAGCG-AATCAAA 704
|||||
Cp 1321 GTCCATGATGCTGCTGCTGCTCCATCACTGAGGCGGAATCAAA 1274
|||||

RESULT 10
LOCUS AA513794 605 bp mRNA EST 20-AUG-1997
DEFINITION nh89c05.r1 NCI-CGAP_Brl.1 Homo sapiens cDNA clone IMAGE:965672 5'
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
sequence.
ACCESSION AA513794
VERSION AA513794.1 GI:2252215
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 606)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636940.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2236 Std Error: 0.00
Seq primer: -28ml3 revl ET from Amersham
High quality sequence stop: 476.
Location/Qualifiers
1..606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:965672"
/clone_lib="NCI-CGAP_Brl.1"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. (The normalized
version of this library is NCI-CGAP_Br2.) Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 188 c 146 g 144 t 2 others
ORIGIN

FEATURES
source

Query Match 25.7%; Score 590; DB 18; Length 606;
Best Local Similarity 98.8%; Pred. No. 0.00e+00;
Matches 600; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Db 1 CCCGACCGCGGAGGC-TTCTCTGGCTGGGTCTGGTCTGAGTCCGTCGTCGCC 59
|||||

QY 135 CCCGACCGCGCGAGGCGCTTCTCTGGCTGGGTGCTGCTGAGCTCCGTCCTCGTGGCGCC 194

Db 60 TCGGATCGGAACGAGCGCCAACTCGACACAGATGCTCTGAACGTTCTTCTCATCATCG 119

QY 195 TCGGATCGGAACGAGCGCCAACTCGACACAGATGCTCTGAACGTTCTTCTCATCATCG 254

Db 120 TGGATGACCTCGGCGCCCTCCCTGGGCTGTTATFGGGATAAGCTGTGAGGTCCCAATA 179

QY 255 TGGATGACCTCGGCGCCCTCCCTGGGCTGTTATFGGGATAAGCTGTGAGGTCCCAATA 314

Db 180 TTGACCACTGGCATCCACAGCCTCTTTCACAAATGCCCTTTGGCGACAGNAGTGT 239

QY 315 TTGACCACTGGCATCCACAGCCTCTTTCACAAATGCCCTTTGGCGACAGCAGTGT 374

Db 240 GCGCCCGAGCGGCTTCTTCTCACTGCGCAGGAGCCTGACACACCGCCGCTGAG 299

QY 375 GCGCCCGAGCGGCTTCTTCTCACTGCGCAGGAGCCTGACACACCGCCGCTGAG 434

Db 300 ACTTCAACTCTACTGAGGGTGCACGCTGGAACCTTCTCCACCATCCCGCAGTACTTCA 359

QY 435 ACTTCAACTCTACTGAGGGTGCACGCTGGAACCTTCTCCACCATCCCGCAGTACTTCA 494

Db 360 AGGAGATGGCTATGTGACCATGTGGGTGGGAAAGTCTTTACCCCTGGGATATCTTCA 419

QY 495 AGGAGATGGCTATGTGACCATGTGGGTGGGAAAGTCTTTACCCCTGGGATATCTTCA 554

Db 420 ACCATACGATGATTCGGTATAGCTGGTCTTTTCCACCTTATCATCTTCTCTGAGA 479

QY 555 ACCATACGATGATTCGGTATAGCTGGTCTTTTCCACCTTATCATCTTCTCTGAGA 614

Db 480 AGTATGAACACATGAAGACATGTCCAGGCGCAGATGGAGACTCCATGCCAACCTGCTTT 539

QY 615 AGTATGAACACATGAAGACATGTCCAGGCGCAGATGGAGACTCCATGCCAACCTGCTTT 674

Db 540 GCCCTGTGGATGTGCTGGATGTTCCGAGGCGACCTTCCCTGACAAACAGACACTGAGC 599

QY 675 GCCCTGTGGATGTGCTGGATGTTCCGAGGCGACCTTCCCTGACAAACAGACACTGAGC 734

Db 600 AAGCCAT 606

QY 735 AAGCCAT 741

RESULT 11

LOCUS AAL131455 644 bp mRNA EST 14-MAY-1997

DEFINITION z132d04.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:503623 5', similar to gb:M58342 IDURONATE 2-SULFATASE

PRECUSOR (HUMAN);, mRNA sequence.

ACCESSION AAL131455

VERSION AAL131455.1 GI:1692996

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 644)

AUTHORS Hallier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissone,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trivasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

COMMENT On Nov 29, 1993 this sequence version replaced gi:503248. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu

This clone is available royalty-free through LLMU ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 849 Std Error: 0.00 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 490.

FEATURES

source

1..644

location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:503623"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAATTCGGCGGCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Patima Bonaldo."

BASE COUNT 147 a 156 c 152 g 183 t 6 others

ORIGIN

Query Match 25.0%; Score 574; DB 13; Length 644;

Best Local Similarity 96.9%; Pred. No. 0.00e+00;

Matches 624; Conservative 0; Mismatches 13; Indels 7; Gaps 7;

Db 1 GGTGGGCTCTAGTGAACATGGGCGCAATACACAAATTTTGATGTGCTACCC 60

QY 1131 GGTGGGCTCTAGTGAACATGGGCGCAATACACAAATTTTGATGTGCTACCC 1190

Db 61 ATGTTCCCTGATATCTATGTTCTTCCGAAAGGAGCGGCTTCACATCCCGAGGCGGAGG 120

QY 1191 ATGTTCCCTGATATCTATGTTCTTCCGAAAGGAGCGGCTTCACATCCCGAGGCGGAGG 1250

Db 121 AGCTTTTCCCTTACCTCGACCCCTTTTGTATTCGGCCTCACAGTTGATGGAGCCAGGCGG 180

QY 1251 AGCTTTTCCCTTACCTCGACCCCTTTTGTATTCGGCCTCACAGTTGATGGAGCCAGGCGG 1310

Db 181 AATCCATGAGCTTGTGGAACCTTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGAC 240

QY 1311 AATCCATGAGCTTGTGGAACCTTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGAC 1370

Db 241 TGCAGGTTCCACATCGCTGCCGCTTCTTCATTTCCAGTTGAGCTGTGAGAGAGGCA 300

QY 1371 TGCAGGTTCCACATCGCTGCCGCTTCTTCATTTCCAGTTGAGCTGTGAGAGAGGCA 1430

Db 301 AGAACCTTCTGAAGCATTTTTCGATTCGTCGCTTGGAGAGGATCCGCTACCTCCCTGGTA 360

QY 1431 AGAACCTTCTGAAGCATTTTCGATTCGTCGCTTGGAGAGGATCCGCTACCTCCCTGGTA 1490

Db 361 ATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCCAGACATCCCTCAGTGA 420

QY 1491 ATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCCAGACATCCCTCAGTGA 1550

Db 421 ATTCTGACAGCGAGTTTAAAGATATAAGATCATGGGCTATTCCTTCAGCCACCATAG 480

QY 1551 ATTCTGACAGCGAGTTTAAAGATATAAGATCATGGGCTATTCCTTCAGCCACCATAG 1610

Db 481 ACTATAGTTACTGTGTGGTTGGCTTCAATCCCTCATGAATTCAGTAACTTTCT 540

QY 1611 ACTATAGTTACTGTGTGGTTGGCTTCAATCC -TGATGAATTTCTAGCTAACTTTCT 1669

Db 541 GACCATTCAGCAGGCGGAGACGTTATTTGGTGGAAATTTCTGACCATTTGCAGGATNCACAA 600

QY 1670 GAC-ATCCATGCAGGCGGAGACTGTATTTTG-TGGA-TTCTGACCATTTGCAGGAT-CACAA 1725

Db 601 TATGTAGNATGAATTTCCCAAGGTTGAAGATCTTTTCNAGTTGTT 644

QY 1726 TATGTATAATGA-TTCCCAAGGTGGA-CATCTTTTCCAGTTGTT 1767

RESULT 12
LOCUS AI478887 629 bp mRNA EST 14-APR-1999
DEFINITION tm25b02.x1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
IMAGE:2157579.3' similar to gb:M58342 IDURONATE 2-SULFATASE
PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AI478887
VERSION AI478887.1 GI:4372055
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 629)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187940.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 812 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 453.
Location/Qualifiers

FEATURES
source
1..629
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2157579"
/clone_lib="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 215 a 101 c 110 g 202 t 1 others
ORIGIN
Query Match 24.5%; Score 563; DB 32; Length 629;
Best Local Similarity 97.8%; Pred. No. 0.00e+00;
Matches 590; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

Db 13 GAAAGAAAACATGGATTACATGTTAAACATAATATTTTGAATATATAGGGCTCATAAGAT 72
|||||
Cp 2296 GAAAGAAAACATGGATTACATGTTAAACATAATATTTTGAATATATAGGGCTCATAAGAT 2237
|||||
Db 73 ATATTATTAATTAACAAAAATAATCAGGAACCTTAGTGTGTCCTAGTTTGGATATAT 132
Cp 2236 ATATTATTAATTAACAAAAATAATCAGGAACCTTAGTGTGTCCTAGTTTGGATATAT 2177
|||||
Db 133 GATTACTTTTGAATGCCTAAATCCACAATAATGAAGATATCTTTGTGTATATG 192
Cp 2176 GATTACTTTTGAATGCCTAAATCCACAATAATGAAGATATCTTTGTGTATATG 2117
|||||
Db 193 TTATGTTTGGTTATATGATGCTTCGTATCCAAAGGTATGACATAACTTTGAGTTGT 252
|||||
Cp 2116 TTATGTTTGGTTATATGATGCTTCGTATCCAAAGGTATGACATAACTTTGAGTTGT 2057
|||||
Db 253 TTGCTTTGTTATTTTACGTAAATAAGCCGTAACCTGTTTAAAAAGAGGAAATTAATAA 312
|||||

Cp 2056 TTGCTTTGTTATTTTATTCAGTAATAAGCCGTAACTGTTTTTAAAAAGAGGAAATTAATAA 1997
Db 313 AAAA-CTGGTCCAATTACCAATTAATAATTTTAATAAGACTAAACAGAGAGTTTGGCT 371
|||||
Cp 1996 AAAAAGTGGTCCAATTACCAATTAATAATTTTAATAAGACTAAACAGGAAAGTTTGGCT 1937
Db 372 GTTACATATTCTCAGGCCAAATTTGTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGC 431
|||||
Cp 1936 GTTACATATTCTCAGGCCAAATTTGTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGC 1877
Db 432 TTCCAATATTATGGGTAAATCACAACAGCAGCTCTAACTCTCTCTCTCACCAGCTGGA 491
|||||
Cp 1876 TTCCAATATTATGGGTAAATCACAACAGCAGCTCTAACTCTCTCTCACCAGCTGGA 1817
Db 492 AGGAGCACATCACATTTTCCCATCCATGTTGGCAAACTCAAGGCATCAACACTCTGGGA 551
|||||
Cp 1816 AGGAGCACATCACATTTTCCCATCCATGTTGGCAAACTCAAGGCATCAACACTCTGGGA 1757
Db 552 AAGATCTTCACCTTGGGATCATTTATACATATTGTTGATC-TGCAATGG-TCAGAAGTCAC 609
|||||
Cp 1756 AAGATCTTCACCTTGGGATCATTTATACATATTGTTGATCCTGCAATGGTCAAGAATCCAC 1697
Db 610 AAA 612
Cp 1696 AAA 1694

RESULT 13
LOCUS AA236900 576 bp mRNA EST 03-MAR-1997
DEFINITION zs43h10.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:688003 3'
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
sequence.
ACCESSION AA236900
VERSION AA236900.1 GI:1860930
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 576)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397523.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -41m13 fwd. Et from Amersham
High quality sequence stop: 306.
Location/Qualifiers

FEATURES
source
1..576
/organism="Homo sapiens"
/db_xref="GDB:5592159"
/db_xref="taxon:9606"
/clone="IMAGE:688003"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBPU, and fetal heart NBHL19W) were mixed, and ss circles

Db 541 AACTCCATGCCAACCTGCTTTGGCCT 566
 ||||||||||||||||||||||||||||||||
 QY 654 AACTCCATGCCAACCTGCTTTGGCCT 679

RESULT 15

LOCUS AI418603 551 bp mRNA EST 09-MAR-1999
 DEFINITION t937d08.x1 Soares_NFL_T_GBC_s1 Homo sapiens cDNA clone
 IMAGE:2110959 3' similar to gb:M58342 IDURONATE 2-SULFATASE
 PRECURSOR (HUMAN); mRNA sequence.

ACCESSION AI418603

VERSION AI418603.1 GI:4264534

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

On Mar 10, 1998 this sequence version replaced gi:2948359.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -400P from Gibco

High quality sequence stop: 453.

FEATURES

source

Location/Qualifiers

1..551

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2110959"

/lab_host="PH10B"

/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NbHL19W, testis NHR, and B-cell

NCI-CGAP_GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 116 a 176 c 128 g 131 t

ORIGIN

Query Match 23.9%; Score 549; DB 31; Length 551;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AATGCGGCACCCGCGGAGCGGCTTCTGTGGTGGGTCTGTTCTGAGCTCCG 60

|||||
 QY 123 AATGCGGCACCCGCGGAGCGGCTTCTGTGGTGGGTCTGTTCTGAGCTCCG 182

|||||
 Db 61 TCTGCGTCGCGCTCGGATCCGAACGAGCGGCTTCTGTGGTGGGTCTGAGCTTC 120

|||||
 QY 183 TCTGCGTCGCGCTCGGATCCGAACGAGCGGCTTCTGTGGTGGGTCTGAGCTTC 242

|||||
 Db 121 TTCTCATCATCGTGAGTACCTGGCGCCCTCCCTGGCTGTTATGGGATGAAGCTGGA 180

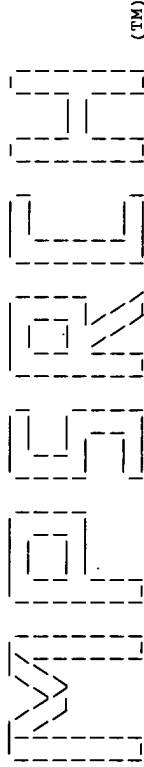
|||||
 QY 243 TTCTCATCATCGTGAGTACCTGGCGCCCTCCCTGGCTGTTATGGGATGAAGCTGGA 302

|||||
 Db 181 GGTCGCCAAATATGACCAATGGCATCCACAGCCCTCTCTCCAGAATGCCCTTTGGCG 240

|||||
 QY 303 GGTCGCCAAATATGACCAATGGCATCCACAGCCCTCTCTCCAGAATGCCCTTTGGCG 362

Db 241 AGCAAGCAGTGTGCGCGCGCGCGGTTTCTTCTCACTGGCAGGAGACCTGACACCA 300
 ||||||||||||||||||||||||||||||||
 QY 363 AGCAAGCAGTGTGCGCGCGCGCGGTTTCTTCTCACTGGCAGGAGACCTGACACCA 422
 ||||||||||||||||||||||||||||||||
 Db 301 CCGGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACCTTCTCCACCATCC 360
 ||||||||||||||||||||||||||||||||
 QY 423 CCGGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACCTTCTCCACCATCC 482
 ||||||||||||||||||||||||||||||||
 Db 361 CCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGTGGGAAAAGTCTTTTCACCCCTG 420
 ||||||||||||||||||||||||||||||||
 QY 483 CCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGTGGGAAAAGTCTTTTCACCCCTG 542
 ||||||||||||||||||||||||||||||||
 Db 421 TGATATCTTCTAACCATACCGGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCATC 480
 ||||||||||||||||||||||||||||||||
 QY 543 GGATATCTTCTAACCATACCGGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCATC 602
 ||||||||||||||||||||||||||||||||
 Db 481 CTTCTCTCTGAGAAATGATGAAAACACTAGACATGTGAGGCGCAGATGAGAACTCCATG 540
 ||||||||||||||||||||||||||||||||
 QY 603 CTTCTCTCTGAGAAATGATGAAAACACTAGACATGTGAGGCGCAGATGAGAACTCCATG 662
 ||||||||||||||||||||||||||||||||
 Db 541 CCAACCTGCTT 551
 ||||||||
 QY 663 CCAACCTGCTT 673

Search completed: Thu Apr 20 18:21:41 2000
 Job time : 7638 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Apr 20 18:22:00 2000; MasPar time 483.16 Seconds
Tabular output not generated. 1150.687 Million cell updates/sec

Title: >US-09-249-003-1
Description: (1-2297) from US09249003.seq
Perfect Score: 2297
N.A. Sequence: 1 CGGCTGTGTCGCCAGTCTT.....TAATCATGTTCTTTTTC 2297
Comp: GCCGACACACGCGTCAGAA.....ATTAGTACAAAGAAAGG

Scoring table: TABLE default
Gap 6
Mismatch STD : Dbase 0; Query 0
Searched: 311545 seqs, 121019393 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: n-geneseq36
1:geneseqn
Statistics: Mean 9.877; Variance 5.713; scale 1.729

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	2297	100.0	2297	1 V15698 Human IDS cDNA.	0.00e+00
2	2297	100.0	2297	1 V52836 Human iduronate-2-sulf	0.00e+00
3	996	43.4	4428	1 V15700 Human IDS gene.	0.00e+00
4	996	43.4	4428	1 V52837 Human iduronate-2-sulf	0.00e+00
5	231	10.1	233	1 Q77329 Human genome fragment	6.01e-140
6	48	2.1	204	1 N81164 Base substituted E.col	1.92e-13
7	48	2.1	204	1 N81164 Base substituted E.col	1.92e-13
8	45	2.0	91	1 Q51746 Oligonucleotide probe	1.24e-11
9	40	1.7	91	1 Q51746 Oligonucleotide probe	1.09e-08
10	37	1.6	501	1 N50033 Sequence encoding new	5.70e-07
11	37	1.6	114955	1 X53491 Human adenosine Al rec	5.70e-07
12	35	1.5	114	1 Q70468 Generic DNA sequence t	7.55e-06
13	34	1.5	114	1 Q70468 Generic DNA sequence t	2.70e-05
14	35	1.5	178	1 T76405 Human endothelin-1 ant	7.55e-06
15	35	1.5	190	1 T76452 Chymase antisense olig	7.55e-06
16	35	1.5	190	1 X54768 Chymase antisense olig	7.55e-06
17	35	1.5	498	1 N50034 Sequence encoding new	7.55e-06
18	35	1.5	501	1 N50023 Sequence encoding new	7.55e-06
19	35	1.5	501	1 N50027 Sequence encoding new	7.55e-06
20	35	1.5	501	1 N50025 Sequence encoding new	7.55e-06

21	35	1.5	501	1 N50032	Sequence encoding new	7.55e-06
22	35	1.5	501	1 N50028	Sequence encoding new	7.55e-06
23	35	1.5	501	1 N50031	Sequence encoding new	7.55e-06
24	35	1.5	530	1 X54739	Human endothelin-1 ant	7.55e-06
25	35	1.5	567	1 X54764	Chymase antisense olig	7.55e-06
26	35	1.5	6225	1 X55273	Human enzyme-related a	7.55e-06
27	33	1.4	91	1 V44650	Mammalian DNA replicat	9.50e-05
28	33	1.4	91	1 V44650	Mammalian DNA replicat	9.50e-05
29	33	1.4	114	1 Q70469	Generic DNA sequence t	9.50e-05
30	32	1.4	114	1 Q70465	Generic DNA sequence t	3.30e-04
31	32	1.4	114	1 Q70470	Generic DNA sequence t	3.30e-04
32	32	1.4	114	1 Q70467	Generic DNA sequence t	3.30e-04
33	32	1.4	114	1 Q70466	Generic DNA sequence t	3.30e-04
34	33	1.4	501	1 N50030	Sequence encoding new	9.50e-05
35	33	1.4	501	1 N50029	Sequence encoding new	9.50e-05
36	33	1.4	7803	1 X55272	Human receptor-related	9.50e-05
37	31	1.3	114	1 Q70465	Generic DNA sequence t	1.13e-03
38	31	1.3	172	1 T76363	Human interleukin 8 an	1.13e-03
39	31	1.3	196	1 X54717	Human IL-8 antisense o	1.13e-03
40	31	1.3	266	1 X54711	Human IL-8 antisense o	1.13e-03
41	31	1.3	501	1 N50024	Sequence encoding new	1.13e-03
42	31	1.3	600	1 V86185	EST clone J635.	1.13e-03
43	31	1.3	600	1 V86185	EST clone J635.	1.13e-03
44	31	1.3	7036	1 X55274	Human factor-related a	1.13e-03
45	31	1.3	114955	1 X53491	Human adenosine Al rec	1.13e-03

ALIGNMENTS

RESULT 1
ID V15698 standard; cDNA; 2297 BP.
AC V15698;
DT 01-JUN-1998 (first entry)
DE Human IDS cDNA.
KW Human; iduronate 2-sulphatase; IDS; treatment;
KW Hunter syndrome; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 125..1777
FT /*tag= a
FT /*product= IDS
PN US5728381-A.
PD 17-MAR-1998.
PF 07-JUN-1995; 484493.
PR 17-DEC-1992; US-991973.
PR 12-NOV-1991; US-790362.
PR 28-NOV-1994; US-345212.
PR 07-JUN-1995; US-484493.
PA (ANSO/) ANSON D S.
PA (BIEL/) BIELICKI J.
PA (CLEM/) CLEMENTS P R.
PA (HOPW/) HOPWOOD J J.
PA (MORR/) MORRIS C P.
PA (OCCH/) OCCHIODORO T.
PA (WILS/) WILSON P J.
PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,
PI Occhiodoro T, Wilson PJ;
DR WPI; 98-206530/18.
DR P-PSDB; W47298.
PT Treatment of iduronate 2-sulphatase deficiency - comprises
PT administering recombinant iduronate 2-sulphatase
PS Example 1; Columns 23-28; 53pp; English.
CC The present sequence encodes human iduronate 2-sulphatase (IDS).
CC IDS deficiency can be treated by administering a recombinant human
CC enzyme, useful in the treatment of Hunter syndrome. The recombinant
CC IDS may be administered in 0.5 microg/kg to 20 mg/kg doses. The
CC administration route is oral, intravenous, intraperitoneal,
CC intramuscular, subcutaneous or intranasal. The recombinant IDS has
CC better uptake properties and/or a longer half-life in vivo, and is
CC thus more efficient than naturally glycosylated IDS.
SQ Sequence 2297 BP; 553 A; 600 C; 510 G; 634 T;

Query Match	100.0%	Score 2297	DB 1	Length 2297
Best Local Similarity 100.0%				
Matches 2297	Conservative	0	Mismatches 0	Indels 0
Gaps				
1	CGGCTGTGTTGGCAGCTCTTCATGGGTTCCGACGAGGAGGTCTCTGTGCTCGCGGCG 60			
1	CGGCTGTGTTGGCAGCTCTTCATGGGTTCCGACGAGGAGGTCTCTGTGCTCGCGGCG 60			
61	TGCTAACTCGGCCACCTGTCTGACGCTGTCCCGCCGCTCTGAAGCGGCGCGTGAAGC 120			
61	TGCTAACTCGGCCACCTGTCTGACGCTGTCCCGCCGCTCTGAAGCGGCGCGTGAAGC 120			
121	CGAAATGCGCCACCCGAGCGGCGAGGCTTCTGTGGCTGGTCTGTGAGCTC 180			
121	CGAAATGCGCCACCCGAGCGGCGAGGCTTCTGTGGCTGGTCTGTGAGCTC 180			
181	CGTCTGCGTGGCCTCGGATCCGAAACGAGGCGCAACTCGACCACACAGATGCTCTGACGT 240			
181	CGTCTGCGTGGCCTCGGATCCGAAACGAGGCGCAACTCGACCACACAGATGCTCTGACGT 240			
241	TCTTCTCATCATCGTGGATGACCTGCGGCCCTCCCTGGGCTGTTATGGGATAGCTGGT 300			
241	TCTTCTCATCATCGTGGATGACCTGCGGCCCTCCCTGGGCTGTTATGGGATAGCTGGT 300			
301	GAGTCCCCAAATATTGACCACTGGCATCCACAGGCTCCTCTTCAGAAATGCTTTGC 360			
301	GAGTCCCCAAATATTGACCACTGGCATCCACAGGCTCCTCTTCAGAAATGCTTTGC 360			
361	GCAGCAAGCAGTGTGCGCCCGGAGCGGTTCTTCTTCCACACGAGGAGACTGACAC 420			
361	GCAGCAAGCAGTGTGCGCCCGGAGCGGTTCTTCTTCCACACGAGGAGACTGACAC 420			
421	CACCGCCTGTAGGACTTCAACTCTACTGGAGGGTGCACGCTGGAACCTTCTCCACCAT 480			
421	CACCGCCTGTAGGACTTCAACTCTACTGGAGGGTGCACGCTGGAACCTTCTCCACCAT 480			
481	CCCCAGTACTCAAGGAGAAATGGCTATGTGACCATGTGCGTGGGAAAAGCTTTCAACC 540			
481	CCCCAGTACTCAAGGAGAAATGGCTATGTGACCATGTGCGTGGGAAAAGCTTTCAACC 540			
541	TGGGATATCTCTAACCATACCGATGATCTCCGATAGCTGGTCTTTTCCACCTTATCA 600			
541	TGGGATATCTCTAACCATACCGATGATCTCCGATAGCTGGTCTTTTCCACCTTATCA 600			
601	TGCTTCTCTGGAAGTATGAAACACTAAGACATGTGAGGGCCAGATGGAGAACTCCA 660			
601	TGCTTCTCTGGAAGTATGAAACACTAAGACATGTGAGGGCCAGATGGAGAACTCCA 660			
661	TGCAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCGAGGGCACCTTGCTGACAA 720			
661	TGCAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCGAGGGCACCTTGCTGACAA 720			
721	ACAGAGACTGACGACGACCATACAGTTGTTGGAAAGATGAAAACGTACCCAGTCCCTT 780			
721	ACAGAGACTGACGACGACCATACAGTTGTTGGAAAGATGAAAACGTACCCAGTCCCTT 780			
781	CTTCTCGGCGTGGGTATCATAAAGCCACACATCCCTTCAGATACCCCAAGGAATTTCA 840			
781	CTTCTCGGCGTGGGTATCATAAAGCCACACATCCCTTCAGATACCCCAAGGAATTTCA 840			
841	GAAGTTGTATCCCTTGGAGAACATACCCCTGGCCCCCGATCCCGAGGTCCCTGATGSCCT 900			
841	GAAGTTGTATCCCTTGGAGAACATACCCCTGGCCCCCGATCCCGAGGTCCCTGATGSCCT 900			
901	ACCCCTGTGGCCTACAAACCCCTGGATGGACATCAGGCAACGGGAGACGTCCAAGCTTT 960			
901	ACCCCTGTGGCCTACAAACCCCTGGATGGACATCAGGCAACGGGAGACGTCCAAGCTTT 960			
961	AAACATCACTGTGCCCTATGTTGCCAATTCCTGTGGACTTTTCAGCGGAAAATCCGCCAG 1020			
961	AAACATCACTGTGCCCTATGTTGCCAATTCCTGTGGACTTTTCAGCGGAAAATCCGCCAG 1020			
1021	CTACTTTGCTCTGTGTCAATTTTGGATACACAGTTCGGCGGCTCTTGTAGTCTTTGGA 1080			

Qy	1021	CTACTTTGGCCTCTGTGTGTCATATTGTGGATACACAGSTGCGCGCGCTTGTAGTGTCTTTGGG	1080
Db	1081	CGATCTTCACGTGGCCAAACAGCACCATCATTCGATTTACCTCGGATCATGGTGGGCTCT	1140
Qy	1081	CGATCTTCACGTGGCCAAACAGCACCATCATTCGATTTACCTCGGATCATGGTGGGCTCT	1140
Db	1141	AGSTGAACATGSGAATGGGCCAAATACAGCAAAATTTTGATGTGCTACCCATGTTCCCT	1200
Qy	1141	AGSTGAACATGSGAATGGGCCAAATACAGCAAAATTTTGATGTGCTACCCATGTTCCCT	1200
Db	1201	GATATCTATGTTCTTGGAAAGACGGCTTCACCTTCGGAGGACGGCGAAGAGCTTTTCCC	1260
Qy	1201	GATATCTATGTTCTTGGAAAGACGGCTTCACCTTCGGAGGACGGCGAAGAGCTTTTCCC	1260
Db	1261	TTACTCGACCCCTTTTGATTTCCGCTCACAGTTGATGGAGCCAGCAGGCAATCCATGGA	1320
Qy	1261	TTACTCGACCCCTTTTGATTTCCGCTCACAGTTGATGGAGCCAGCAGGCAATCCATGGA	1320
Db	1321	CCTTGTGGAACCTTGTGTCTCTTTTCCCACGCTGGCTGAGACTTGCAGGACTGCAAGTTCC	1380
Qy	1321	CCTTGTGGAACCTTGTGTCTCTTTTCCCACGCTGGCTGAGACTTGCAGGACTGCAAGTTCC	1380
Db	1381	ACCTCGCTGCCCTTCTTTCATTTACGTTTGAGCTGTGCTAGAGAAAGCAAGAACTTCT	1440
Qy	1381	ACCTCGCTGCCCTTCTTTCATTTACGTTTGAGCTGTGCTAGAGAAAGCAAGAACTTCT	1440
Db	1441	GAAGCATTTTCGATTCGCTGACTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA	1500
Qy	1441	GAAGCATTTTCGATTCGCTGACTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA	1500
Db	1501	ACTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGCACAA	1560
Qy	1501	ACTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGCACAA	1560
Db	1561	CGCGAGTTTAAAAGATATAAAGATCATGGGCTATTCCATACGCACCATAGACTATAGTA	1620
Qy	1561	CGCGAGTTTAAAAGATATAAAGATCATGGGCTATTCCATAGCCACCATAGACTATAGTA	1620
Db	1621	TACTGTGTGGGTTGGCTTCAATCCTCATCAATTTCTAGTAACTTTCTGCATCCATGTC	1680
Qy	1621	TACTGTGTGGGTTGGCTTCAATCCTCATCAATTTCTAGTAACTTTCTGCATCCATGTC	1680
Db	1681	AGGGAACTGATTTTGTGGATTTCTGACCCATGCGAGGATCACAAATGTAATGATTC	1740
Qy	1681	AGGGAACTGATTTTGTGGATTTCTGACCCATGCGAGGATCACAAATGTAATGATTC	1740
Db	1741	CCAAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTTTGGCAACCATGGATGGCAA	1800
Qy	1741	CCAAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTTTGGCAACCATGGATGGCAA	1800
Db	1801	ATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTCTTTTGTGATTT	1860
Qy	1801	ATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTCTTTTGTGATTT	1860
Db	1861	ACCCATAATATTGGAAGACGCTTGAGGCTAGTAAATCCAAACATGCATCAACAATTTGG	1920
Qy	1861	ACCCATAATATTGGAAGACGCTTGAGGCTAGTAAATCCAAACATGCATCAACAATTTGG	1920
Db	1921	CCTGAGAAATATGTAACAGCCAAACCTTTTTCGTTTATGCTTTTATTAATAATTTAATTTGGT	1980
Qy	1921	CCTGAGAAATATGTAACAGCCAAACCTTTTTCGTTTATGCTTTTATTAATAATTTAATTTGGT	1980
Db	1981	AATTTGGACCACTTTTATTTTAAATTCCTCTTTTAAACAGTTACGGCTTATTTACTG	2040
Qy	1981	AATTTGGACCACTTTTATTTTAAATTCCTCTTTTAAACAGTTACGGCTTATTTACTG	2040
Db	2041	AATAAATACAAAGCAACAAACTCAAGTTATGTGCATACCTTTGGATACGAAGACCATACA	2100
Qy	2041	AATAAATACAAAGCAACAAACTCAAGTTATGTGCATACCTTTGGATACGAAGACCATACA	2100
Db	2101	TAATAACCAACATAACATTTATACACAAAGAACTTTTCATTTATTTGGTAATTTAGTGC	2160

Qy	121	CGAAATCGCCGACCCCGGACCGCCGAGGCGCTTCTCTGGCTGGGTCTGTGGTTCTGAGTC	180
Db	181	CGTCTGGTCCGCTCGGATCCGAAACGAGCGCAACTCGACACAGATGTCTTGAAGT	240
Qy	181	CGTCTGGTCCGCTCGGATCCGAAACGAGCGCAACTCGACACAGATGTCTTGAAGT	240
Db	241	TCCTTCATCATCGTGGATGACCTGGCCCTTCCCTGGGCTGTTATGGGGATAAGCTGGT	300
Qy	241	TCCTTCATCATCGTGGATGACCTGGCCCTTCCCTGGGCTGTTATGGGGATAAGCTGGT	300
Db	301	GAGTCCCCAAATATTGACCACTGCATCCCAACAGCCCTCTTCTTCCAGAATGCTTTGC	360
Qy	301	GAGTCCCCAAATATTGACCACTGCATCCCAACAGCCCTCTTCTTCCAGAATGCTTTGC	360
Db	361	GCAGCAAGCAGTGTGGCCCGGAGCGGCTTCTTCTTCTCACTGGCAGGAGACCTGACAC	420
Qy	361	GCAGCAAGCAGTGTGGCCCGGAGCGGCTTCTTCTTCTCACTGGCAGGAGACCTGACAC	420
Db	421	CACCCGCTTACGACTTCAACTCTTACTGGAGGTGCACGCTGGAATCTTCTCCACAT	480
Qy	421	CACCCGCTTACGACTTCAACTCTTACTGGAGGTGCACGCTGGAATCTTCTCCACCAT	480
Db	481	CCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTGGTGGGAAAGTCTTTCACCC	540
Qy	481	CCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTGGTGGGAAAGTCTTTCACCC	540
Db	541	TGGGATATCTTCTAAACCATACCGATCATCTCCGTATAGCTCGTCTTCTTCCACCTTATCA	600
Qy	541	TGGGATATCTTCTAAACCATACCGATCATCTCCGTATAGCTCGTCTTCTTCCACCTTATCA	600
Db	601	TCCTTCCTCTGAGAAGTATGAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCA	660
Qy	601	TCCTTCCTCTGAGAAGTATGAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCA	660
Db	661	TGCCAAGCTGTTTGCCCTGTGGATGTGCTGGAGTFTCCCGAGGGACCTTGCCTGACAA	720
Qy	661	TGCCAAGCTGTTTGCCCTGTGGATGTGCTGGAGTFTTCCCGAGGGACCTTGCCTGACAA	720
Db	721	ACAGAGCAGTGCAGAGCCATACAGTGTGTGGAAAGATGAAAGCTCAGCAGTCCCTTT	780
Qy	721	ACAGAGCAGTGCAGAGCCATACAGTGTGTGGAAAGATGAAAGCTCAGCAGTCCCTTT	780
Db	781	CTTCTCGCCGTTGGGTATCATAAAGCCACACATCCCTTCAGATACCCCAAGAAATTTCA	840
Qy	781	CTTCTCGCCGTTGGGTATCATAAAGCCACACATCCCTTCAGATACCCCAAGAAATTTCA	840
Db	841	GAGTTGTATCCCTTGGAGAACATACCTTGGCCCCCGATCCCGAGGTCCCTGTATGGCCT	900
Qy	841	GAGTTGTATCCCTTGGAGAACATACCTTGGCCCCCGATCCCGAGGTCCCTGTATGGCCT	900
Db	901	ACCCCTGTGGCTTCAACCCCTGGATGACATCAGGCAACGGGAAGAGCTCCAGCCCTT	960
Qy	901	ACCCCTGTGGCTTCAACCCCTGGATGACATCAGGCAACGGGAAGAGCTCCAGCCCTT	960
Db	961	AAACATCAGTGTGCCCTATGGTCCAATTCTGTGGACTTTCAGCGGAAATCCCGCAGAG	1020
Qy	961	AAACATCAGTGTGCCCTATGGTCCAATTCTGTGGACTTTCAGCGGAAATCCCGCAGAG	1020
Db	1021	CTACTTTGCCCTGTGTGCATATTTGGATACACAGGTGGCGGCCCTTCTGTAGTGTCTTGA	1080
Qy	1021	CTACTTTGCCCTGTGTGCATATTTGGATACACAGGTGGCGGCCCTTCTGTAGTGTCTTGA	1080
Db	1081	CGATCTTCAGCTGGCCAAACAGCACCATCATTCATTTACCTCGGATCATGGTGGGCTCT	1140
Qy	1081	CGATCTTCAGCTGGCCAAACAGCACCATCATTCATTTACCTCGGATCATGGTGGGCTCT	1140
Db	1141	AGGTGAACATGGAGAATGGGCCAAATPACAGCAATTTTGATGTGTCTACCCATGTTTCCCT	1200
Qy	1141	AGGTGAACATGGAGAATGGGCCAAATPACAGCAATTTTGATGTGTCTACCCATGTTTCCCT	1200
Db	1201	GATATCTATGTCTTGGAGGACGGCTTCACTTCCGGAGGACGCGGAGAAAGCTTTTCCC	1260
Qy	1201	GATATCTATGTCTTGGAGGACGGCTTCACTTCCGGAGGACGCGGAGAAAGCTTTTCCC	1260

```
Db 1261 TTACCTCGACCCCTTTTGATTTCCGGCTCACAGTTGATGGAGCCAGGCAATCCATGGA 1320
QY 1261 TTACCTCGACCCCTTTTGATTTCCGGCTCACAGTTGATGGAGCCAGGCAATCCATGGA 1320
Db 1321 CCTTGTTGGAACCTGTGCTCTTTTCCACCGCTGGCTGGACTTCGACGACTGCAGGTTCC 1380
QY 1321 CCTTGTTGGAACCTGTGCTCTTTTCCACCGCTGGCTGGACTTCGACGACTGCAGGTTCC 1380
Db 1381 ACCTCGCTGCCCTTCCCTTCATTTACGTTTGAGCTGTGCAGAGAAGCAAGAACCTTCT 1440
QY 1381 ACCTCGCTGCCCTTCCCTTCATTTACGTTTGAGCTGTGCAGAGAAGCAAGAACCTTCT 1440
Db 1441 GAAGCATTTTCGATTCGGTGACTTCGTTGGAAGAGGATCCGTAACCTCCCGTGA 1500
QY 1441 GAAGCATTTTCGATTCGGTGACTTCGTTGGAAGAGGATCCGTAACCTCCCGTGA 1500
Db 1501 ACTGATTTGCCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATTCGACAA 1560
QY 1501 ACTGATTTGCCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATTCGACAA 1560
Db 1561 GCCGAGTTTAAAGATATAAAGATCATGGGCTATTCCATAGCGACCATAGACTATAGTA 1620
QY 1561 GCCGAGTTTAAAGATATAAAGATCATGGGCTATTCCATAGCGACCATAGACTATAGTA 1620
Db 1621 TACTGTGTGGGTTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGCATCCCATGC 1680
QY 1621 TACTGTGTGGGTTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGCATCCCATGC 1680
Db 1681 AGGGAACTGTATTTTGTGGATTCGACCCATTCGAGGATCACAATATGTAATGATTC 1740
QY 1681 AGGGAACTGTATTTTGTGGATTCGACCCATTCGAGGATCACAATATGTAATGATTC 1740
Db 1741 CCAAGTGGAGACTTTTCCAGTTGTTGATGCCCTTGAGTTTGGCAACCATGGATGGCAA 1800
QY 1741 CCAAGTGGAGACTTTTCCAGTTGTTGATGCCCTTGAGTTTGGCAACCATGGATGGCAA 1800
Db 1801 ATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTGGTGGTGGATT 1860
QY 1801 ATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTGGTGGTGGATT 1860
Db 1861 ACCCATAAATTTGGAAGCAGCCTGAGGCTAGTTTAAATCCAAACATGCATCAACAAATTTGG 1920
QY 1861 ACCCATAAATTTGGAAGCAGCCTGAGGCTAGTTTAAATCCAAACATGCATCAACAAATTTGG 1920
Db 1921 CCTGAGAAATGTAAACAGCCAAACCTTTTCGTTTAGTCTTTTATTAATTTATATTTGGT 1980
QY 1921 CCTGAGAAATGTAAACAGCCAAACCTTTTCGTTTAGTCTTTTATTAATTTATATTTGGT 1980
Db 1981 AATTGGACCAAGTTTATTTTAAATTTCCCTCTTTTAAATTTAAACAGTTACGGCTTATTTACTG 2040
QY 1981 AATTGGACCAAGTTTATTTTAAATTTCCCTCTTTTAAATTTAAACAGTTACGGCTTATTTACTG 2040
Db 2041 AATAAATACAAAGCAACAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100
QY 2041 AATAAATACAAAGCAACAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100
Db 2101 TAATAACCAACATACATTTATACAAAGAAATCTTTTCATTTATTTGGAAATTTAGTGC 2160
QY 2101 TAATAACCAACATACATTTATACAAAGAAATCTTTTCATTTATTTGGAAATTTAGTGC 2160
Db 2161 ATTTCAAAAAGTATCATATATCAAACTAGGCACCACTAAGTTCCTGATTTATTTGTT 2220
QY 2161 ATTTCAAAAAGTATCATATATCAAACTAGGCACCACTAAGTTCCTGATTTATTTGTT 2220
Db 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATCAAAATATTTATGTTAAACATGAA 2280
QY 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATCAAAATATTTATGTTAAACATGAA 2280
Db 2281 TCCATGTTTCTTTTCC 2297
QY 2281 TCCATGTTTCTTTTCC 2297
```

```
RESULT 3
ID V15700 standard; DNA; 4428 BP.
AC V15700;
DT 01-JUN-1998 (first entry)
DE Human IDS gene.
KW Iduronate 2-sulphatase; IDS; treatment;
KW Hunter syndrome; ds.
OS Homo sapiens.
PN US5728381-A.
PD 17-MAR-1998.
PF 07-JUN-1995; 484493.
PR 17-DEC-1992; US-991973.
PR 12-NOV-1991; US-790362.
PR 28-NOV-1994; US-345212.
PR 07-JUN-1995; US-484493.
PA (ANSO/) ANSON D S.
PA (BIEL/) BIELICKI J.
PA (CLEM/) CLEMENTS P R.
PA (HOPW/) HOPWOOD J J.
PA (MORR/) MORRIS C P.
PA (OCCH/) OCCHIODORO T.
PA (WILS/) WILSON P J.
PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,
Occhiodoro T, Wilson PJ;
WPI; 98-206530/18.
PT Treatment of iduronate 2-sulphatase deficiency - comprises
administering recombinant iduronate 2-sulphatase
Example 1; Columns 33-38; 53pp: English.
CC The present sequence is the human iduronate 2-sulphatase (IDS)
gene. IDS deficiency can be treated by administering a recombinant
human IDS that is more highly glycosylated than the naturally
occurring enzyme, useful in the treatment of Hunter syndrome. The
recombinant IDS may be administered in 0.5 microg/kg to 20 mg/kg
doses. The administration route is oral, intravenous,
intraperitoneal, intramuscular, subcutaneous or intranasal. The
recombinant IDS has better uptake properties and/or a longer
half-life in vivo, and is thus more efficient than naturally
occurring glycosylated IDS.
CC glycosylated IDS.
SQ Sequence 4428 BP; 1033 A; 1064 C; 1042 G; 1285 T;
```

```
Query Match 43.4%; Score 996; DB 1; Length 4428;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3433 CAGGAGGCAATCCATGGACCTTGGGAACCTTGTGCTCTTTTCCACGCTGGCTGGAC 3492
QY 1302 CAGGAGGCAATCCATGGACCTTGGGAACCTTGTGCTCTTTTCCACGCTGGCTGGAC 1361
Db 3493 TTGCAGGACTGCAGGTTCCACCTCGCTGCCCGTTCCCTTCATTTACGTTGAGCTGCA 3552
QY 1362 TTGCAGGACTGCAGGTTCCACCTCGCTGCCCGTTCCCTTCATTTACGTTGAGCTGCA 1421
Db 3553 GAGAAGGCAAGAACCTTCTGGAAGCATTTTCGATTTCCGTCGACTTGGAGAGGATCCGTACC 3612
QY 1422 GAGAAGGCAAGAACCTTCTGGAAGCATTTTCGATTTCCGTCGACTTGGAGAGGATCCGTACC 1481
Db 3613 TCCCTGGTAATCCCGTGAACCTTGCCTATAGCCAGTATCCCGGCTTCAGACATCC 3672
QY 1482 TCCCTGGTAATCCCGTGAACCTTGCCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
Db 3673 CTCAGTGGAAATTCGACAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATGAC 3732
QY 1542 CTCAGTGGAAATTCGACAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATGAC 1601
Db 3733 GCACCATAGACTATAGGTATCTGTGGTTGGCTTCAATCTGATGAATTTCTAGCTA 3792
QY 1602 GCACCATAGACTATAGGTATCTGTGGTTGGCTTCAATCTGATGAATTTCTAGCTA 1661
Db 3793 ACTTTTCTGACATCCATGCAGGGGAACCTATTTTGTGGATTTCTGACCCATTCGAGGATC 3852
QY 1662 ACTTTTCTGACATCCATGCAGGGGAACCTATTTTGTGGATTTCTGACCCATTCGAGGATC 1721
```



```
QY 1362 TTGCAGGACTGCAGGTTCCACCTCGCTGCCCGTTCCTTCATTTACAGTTGAGCTGTGCA 1421
Db GAGAGGCAAGAACTTCTGAAGCATTTTCGATTCGCGTGAAGGAGGATCGGTACC 3612
QY 1422 GAGAGGCAAGAACTTCTGAAGCATTTTCGATTCGCGTGAAGGAGGATCGGTACC 1481
Db TCCTGTGTAATCCCGTGAACCTGATGCTATAGCCAGTATCCCGGCTTCAGACATCC 3672
QY 1482 TCCTGTGTAATCCCGTGAACCTGATGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
Db CTGAGTGAATCTGACAGCCGAGTTTAAAGATATAAAGATATGAAGATATGCGGCTTCATCCATAC 3732
QY 1542 CTCAGTGAATCTGACAGCCGAGTTTAAAGATATAAAGATATGAAGATATGCGGCTTCATCCATAC 1601
Db GCACCATAGACTATAGTATGATCTGCTGGGTGGCTCAATCCTCATCAATTTCTAGCTA 3792
QY 1602 GCACCATAGACTATAGTATGATCTGCTGGGTGGCTCAATCCTCATCAATTTCTAGCTA 1661
Db ACTTTCTGACATCCATGCGAGGGAACCTGATTTTGTGGATCTTGACCATGTCAGGATC 3852
QY 1662 ACTTTCTGACATCCATGCGAGGGAACCTGATTTTGTGGATCTTGACCATGTCAGGATC 1721
Db ACAATATGTAATGATTTCCAGGTGAGATCTTTTCCAGTGTGTGATGCTTGGATTT 3912
QY 1722 ACAATATGTAATGATTTCCAGGTGAGATCTTTTCCAGTGTGTGATGCTTGGATTT 1781
Db TGCCAAACCATGGATGGCAAAATGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 3972
QY 1782 TGCCAAACCATGGATGGCAAAATGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 1841
Db AGCTGGTGGCTTTTCTGATATACCATATATATTTGGAAGCAGCTGAGGGCTAGTTAATCAA 4032
QY 1842 AGCTGGTGGCTTTTCTGATATACCATATATATTTGGAAGCAGCTGAGGGCTAGTTAATCAA 1901
Db ACATGTCATCAACAAATTTGGCCCTGAGATATGTAACAGCAAAACCTTTTCGTTTACTCTTT 4092
QY 1902 ACATGTCATCAACAAATTTGGCCCTGAGATATGTAACAGCAAAACCTTTTCGTTTACTCTTT 1961
Db ATTAAATTTAATATTTGGTAAATGACACAGTGTGTTTTTTTAAATTTCCCTCTTTTAAAC 4152
QY 1962 ATTAAATTTAATATTTGGTAAATGACACAGTGTGTTTTTTTAAATTTCCCTCTTTTAAAC 2021
Db AGTTACGCTTATTTACTGTAATATCAAGCAAAACCAACCAATCAAGTTATGTCATACCTT 4212
QY 2022 AGTTACGCTTATTTACTGTAATATCAAGCAAAACCAACCAATCAAGTTATGTCATACCTT 2081
Db TGGATACGAAGACCATACATATTAACCAACCAATCAACATATATACAAAGATACCTTTCAT 4272
QY 2082 TGGATACGAAGACCATACATATTAACCAACCAATCAACATATATACAAAGATACCTTTCAT 2141
Db TATTTGCGAATTTAGTGCAATTTCAAAAGTAAATCATATATCAAACTAGGCCACCACTA 4332
QY 2142 TATTTGCGAATTTAGTGCAATTTCAAAAGTAAATCATATATCAAACTAGGCCACCACTA 2201
Db AGTTCCCTGATTTATTTGTTTAAATTTAATATATATATCTTATGAGCCCTATATCAAA 4392
QY 2202 AGTTCCCTGATTTATTTGTTTAAATTTAATATATATATCTTATGAGCCCTATATCAAA 2261
Db ATATTATGTTAAACATGTAATCCATGTTCTTTTCC 4428
QY 2262 ATATTATGTTAAACATGTAATCCATGTTCTTTTCC 2297
```

RESULT 5

```
ID 07329 standard; DNA; 233 BP.
AC 07329;
DT 15-NOV-1994 (first entry)
DE Human genome fragment (Preferred).
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN W09401548-A.
PD 20-JAN-1994.
```

```
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield RM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
PT WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 470; 616pp; English.
CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SQ Sequence 233 BP; 77 A; 22 C; 38 G; 96 T;

Query Match 10.1%; Score 231; DB 1; Length 233;
Best Local Similarity 99.6%; Pred. No. 6.01e-140;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 TTAACATAATATTTGAATATATAGGGCTCATAGATATATTAAATTTATAACAAAA 60
Cp 2273 TTAACATAATATTTGAATATATAGGGCTCATAGATATATTAAATTTATAACAAAA 2214
Db 61 TTAACAGGAACCTAGTGTGCTGCTAGTTGATATATGATGATGCTTTTGAATGCACTAA 120
Cp 2213 TTAACAGGAACCTAGTGTGCTGCTAGTTGATATATGATGATGCTTTTGAATGCACTAA 2154
Db 121 ATTCCACAATATGTAAGTATCTTTGCTATGTAAGTATGTTGTTATGTTATGTTATG 180
Cp 2153 ATTCCACAATATGTAAGTATCTTTGCTATGTAAGTATGTTGTTATGTTATGTTATG 2094
Db 181 TCTTCGTATCCAAAGTATGACATAACTTGTGAGTTTCTTGTCTGTTGTTATTT 233
Cp 2093 TCTTCGTATCCAAAGTATGACATAACTTGTGAGTTTCTTGTCTGTTGTTATTT 2041
```

RESULT 6

```
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT primer_bind /tag= a
FT primer_bind /function= multiple cloning site
FT primer_bind 187..204
FT primer_bind /tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI; 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
```


New antisense oligonucleotides used in treatment of, e.g. pulmonary vasoconstriction

PT disclosure; Page 37; 120pp; English.

The specification describes antisense oligonucleotides (X52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences X55272-74. These multiple target oligonucleotides (specifically X55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.

Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match 1.6%; Score 37; DB 1; Length 114955;
Best Local Similarity 34.9%; Pred. No. 5.70e-07;
Matches 51; Conservative 38; Mismatches 57; Indels 0; Gaps 0;

D_b 89952 TGGBGCBRBBBGGCGBGCGGGGGRBGBBBGBBCGCCTCCCBGBGGC 90011
|||| :||| : || : | || : || : || : || : || : || : || : || : ||

Q_y 23 TGGGTCCCACAGGAGGACTCTGTGGCTCGCGCGTCACTAATGCCCATCTGCTGC 82

D_b 90012 TGCTTCGCBBTCTTGTTTCTTTCCRGCTCTTGTTTTTBTCCTCCBGBGGCB 90071
||||| :||| : ||| : | ||| : ||| : ||| : ||| : ||| : ||| :

Q_y 83 AGCGTFCCTCCCGCGCTCTGAAGCGCGCGCTGAAGCGCAAATGCCGCCACCCCGACC 142

D_b 90072 GBGEGGGCBGGNNNNGCGTCT 90097
| : | ||| : | |||

Q_y 143 GGCCGAGCCTCTCTGGCTGGTCT 168

RESULT 12

ID ID Q70468 standard; DNA; 114 BP.
DC O70468;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
sequence of 6, 9 or 12 nucleotides (see comments)"

FN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; US00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA {OINC}- UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK.
DR WPI; 94-279739/34.
DR p-PDSB; R65154.
PT Identifying proteins or peptide(s) which bind a ligand - by
screening a recombinant vector library expressing fusion proteins
comprising a binding domain and an effector domain
PS disclosure; page 35; 255pp; English.
CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
Synthetic Affinity Reagents) peptides This generic formula can also be

CC represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y.X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compns. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 CC Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 1.5%; Score 35; DB 1; Length 114;
 Best Local Similarity 3.7%; Pred. No. 7.55e-06;
 Matches 4; Conservative 33; Mismatches 72; Indels 0; Gaps 0;

Db 6 BNNBNNBNNBNNBNNBNNBNTGCGNNBNNBNNBNNBNNBNNBNNBNNBNN 65
 QY 415 TGACACACCCGCGCTGTACGACTTCAACTCTCTACGGAGGTGCGAACTCTC 474
 Db 66 BNNBNNBNNBNNBNNBNNBNTGCGNNBNNBNNBNNBNNBNNBNNBNNB 114
 QY 475 CACCATCCCCAGTACTTCAAGGAGAATGGCTATGTGACCATGTCGGTG 523

RESULT 13
 ID Q70466 standard; DNA; 114 BP.
 AC Q70466;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR-9 peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FS Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 FT
 PN W09418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UINC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB; R65152.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70466 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)4Z(NNB)8(TGC)(NNB)
 CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
 CC that are not specified further. Other generic sequences are shown in
 CC Q70466-68. Other specific peptides generated by these generic sequences
 CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
 CC or peptides, comprising at least two functional regions - a binding
 CC domain with affinity for a ligand and a second effector peptide portion

CC that is chemically or biologically active. They may further comprise a
 CC linker peptide between the 2 domains. The oligonucleotides are also
 CC designed so that the expressed peptide contains 2 or 4 cysteine residues
 CC positioned in, or flanking, the unpredicted or variant residues. These
 CC residues confer some degree of conformational rigidity to the peptides.
 CC The TSARs or compns. comprising a TSAR binding domain can be used in
 CC vivo to deliver a chemically or biologically active moiety, eg. metal
 CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
 CC on the cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need for
 CC complex methods of hybridoma formation or in vivo antibody production.
 CC The TSARs are easily characterised and have designed activity allowing
 CC direct and rapid detection in a screening process.
 CC Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 1.5%; Score 34; DB 1; Length 114;
 Best Local Similarity 5.7%; Pred. No. 2.70e-05;
 Matches 6; Conservative 31; Mismatches 69; Indels 0; Gaps 0;

Db 9 BNNBNNBNNBNNBNNBNNBNTGCTGCNNBNNBNNBNNBNNBNNBNNBNN 68
 QY 415 TGACACACCCGCGCTGTACGACTTCAACTCTCTACGGAGGTGCGAACTTCTC 474
 Db 69 BNNBNNBNNBNNBNNBNNBNTGCGNNBNNBNNBNNBNNBNNBNNBNNB 114
 QY 475 CACCATCCCCAGTACTTCAAGGAGAATGGCTATGTGACCATGTCG 520

RESULT 14
 ID T76405 standard; DNA; 178 BP.
 AC T76405;
 DT 15-SEP-1997 (first entry)
 DE Human endothelin-1 antisense oligonucleotide.
 KW Asthma; airway epithelium; adenosine free; cystic fibrosis;
 KW chronic obstructive pulmonary disease; bronchitis; ss.
 OS Synthetic.
 PN W09640162-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09306.
 PR 07-JUN-1995; US-474497.
 PA (UYEC-) UNIV EAST CAROLINA.
 PI Metzger WJ NYCE JW;
 DR WPI: 97-051871/05.
 PT Treatment of airway diseases such as asthma - by topically applying
 PT adenosine-free antisense oligo:nucleotide to airway epithelium of
 PT subject
 PS Claim 5; Page 38; 71pp; English.
 CC A method for treating airway disease in a subject has been produced,
 CC which involves the topical administration of an essentially adenosine
 CC free antisense oligonucleotide (ON) to the airway epithelium of the
 CC subject. The present sequence is an antisense oligonucleotide specific
 CC for the human endothelin-1, targeted at the initiation codon. The
 CC method can be used to treat airway diseases such as cystic fibrosis,
 CC asthma, chronic obstructive pulmonary disease, bronchitis and other
 CC airway diseases characterised by an inflammatory response. By
 CC eliminating adenosine from the antisense ON, its liberation upon
 CC antisense degradation is prevented, thereby preventing adenosine-
 CC induced bronchoconstriction in patients with hyper-reactive airways.
 CC Sequence 178 BP; 0 A; 52 C; 46 G; 32 T;

Query Match 1.5%; Score 35; DB 1; Length 178;
 Best Local Similarity 30.5%; Pred. No. 7.55e-06;
 Matches 36; Conservative 43; Mismatches 38; Indels 1; Gaps 1;

Db 53 BCCBCGCGCGCTGCTCBGCGCTBBGCTGCTGTCTTCTGGBGCTCCTTGGCBGCGC 112
 QY GCCTCCTTCTGCTCAGCCTCTCCCGCGCT-CTGAAGCGCGCGCTCGAAGCGAAATG 127
 Db 113 BBBCBGCGBGCBGBBCTGTGCBGBBTTBCTCCBTCTGBBBBSGGGTCTCBBB 170
 QY CCGCCACCCCGGACCGCGCGGCGCTCTCTGCTGGGTCTGTTCTGAGCTCGCTCT 185

RESULT	15
ID	T76452 standard; DNA; 190 BP.
AC	T76452;
AD	16-SBP-1997 (first entry)
DE	Chymase antisenase oligonucleotide.
KW	Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW	chronic obstructive pulmonary disease; bronchitis; ss.
OS	Synthetic.
PN	WO9640162-A1.
PN	19-DEC-1996.
PF	06-JUN-1986; UO9306.
PR	07-JUN-1995; US-474497.
PA	(UYEC-) UNIV EAST CAROLINA.
PI	Metzger WJ, Nyce JW;
DR	WPI; 97-051871/05.
PT	Treatment of airway diseases such as asthma - by topically applying
PT	adenosine-free antisenase oligo-nucleotide to airway epithelium of
PT	subject
PS	Example 5; Page 40; 71pp; English.
CC	A method for treating airway disease in a subject has been produced,
CC	which involves the topical administration of an essentially adenosine
CC	free antisenase oligonucleotide (ON) to the airway epithelium of the
CC	subject. The present sequence is an antisenase oligonucleotide specific
CC	for chymase, targeted at the initiation codon. The method can be
CC	used to treat airway diseases such as cystic fibrosis, asthma, chronic
CC	obstructive pulmonary disease, bronchitis and other airway diseases
CC	characterised by an inflammatory response. By eliminating adenosine
CC	from the antisenase ON, its liberation upon antisenase degradation is
CC	prevented, thereby preventing adenosine-induced bronchoconstriction
CC	in patients with hyper-reactive airways.
SC	Sequence 190 BP; 1 A; 41 C; 58 G; 36 T;
Query Match	1.58; Score 35; DB 1; Length 190;
Best Local Similarity	35.1%; Pred. No. 7.5se+06;
Matches	46; Conservative 37; Mismatches 48; Indels 0; Gaps 0;
Db	58 TGBGBCBGBGBBBBGGCGBGCGGGGGGBGBGBGCBGCTTCCCBGBGBGC 117 :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
Qy	23 TGGETCCCCAGCAGGAGGTCTGTGGCTGCGCGCGGTCAACTGCGCCACTCTGC 82 :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
Db	118 TGCTGTGCBGBBTGCTGGTTTCCTTTCCBGTCTTGCGTTTTTBBCCTCCCBBGBGCCB 177 :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
Qy	83 AGCTGTCCCCGGCGCTCTGAAGCGCGCGGTGAAGCCGAATGCGCCACC CGGACC 142 :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
Db	178 GBGBGGGGBB 188 !:: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
Qy	143 GGCGAGGCGCT 153 :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :

Search completed: Thu Apr 20 18:34:48 2000
Job time : 768 secs.

WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 20 12:57:15 2000; MasPar time 5060.39 Seconds
Tabular output not generated. 1476.553 Million cell updates/sec

Title: >US-09-249-003-1
Description: (1-2297) from US09249003.seq
Perfect Score: 2297

N.A. Sequence: 1 CGGCTGTGTTGGCAGTCTT.....TAATCCAGTTTCTTTTCC 2297
Comp: GCCGACACACGGCGTCAGAA.....ATTAGTACAAAGAAAAGG

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 197718 seqs, 1626453718 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb161

1:em_ba1 2:em_ba2 3:em_fun 4:em_htg1 5:em_htg2 6:em_htg3
7:em_hum1 8:em_hum2 9:em_hum3 10:em_hum4 11:em_in
12:em_om 13:em_or 14:em_ov 15:em_pat 16:em_pl 17:em_ro
18:em_un 19:em_vl

Database:

genbank1-115
20:gb_ba1 21:gb_ba2 22:gb_htg1 23:gb_htg2 24:gb_htg3
25:gb_htg4 26:gb_htg5 27:gb_htg6 28:gb_htg7 29:gb_in1
30:gb_in2 31:gb_om 32:gb_ov

Statistics: Mean 13.051; Variance 5.508; scale 2.369

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
C 1	373	16.2	13574 24	AC011389 Homo sapiens chromosom	0.00e+00
C 2	367	16.0	13574 24	AC011389 Homo sapiens chromosom	0.00e+00
C 3	75	3.3	16310 26	AC015390 Drosophila melanogaste	3.22e-37
C 4	75	3.3	113816 23	AC004358 Drosophila melanogaste	3.22e-37
C 5	40	1.7	1056 31	MVU87256 Mustela vison cr dinuc	2.73e-09
C 6	40	1.7	10772 30	AF012089 Drosophila melanogaste	2.73e-09
C 7	38	1.7	10772 30	AF012089 Drosophila melanogaste	2.73e-09
C 8	37	1.6	1056 31	MVU87256 Mustela vison cr dinuc	3.36e-07
C 9	34	1.5	56722 27	AC016207 Homo sapiens clone RP1	3.42e-05
C 10	34	1.5	56722 27	AC016207 Homo sapiens clone RP1	3.42e-05
C 11	34	1.5	59006 27	AC016220 Homo sapiens clone RP1	3.42e-05
C 12	35	1.5	67656 27	AC016439 Homo sapiens clone RP1	7.50e-06
C 13	34	1.5	176224 22	HSAL175010 Homo sapiens chromosom	3.42e-05

14	34	1.5	190839 23	AL133350 Homo sapiens chromosom	3.42e-05
15	34	1.5	192798 26	AC015864 Homo sapiens chromosom	3.42e-05
C 16	32	1.4	121485 22	HSDJ812P3 Homo sapiens chromosom	6.60e-04
C 17	32	1.4	152878 29	CEY18D10A Caenorhabditis elegans	6.60e-04
C 18	32	1.4	164587 24	AC009444 Homo sapiens clone NHO	6.60e-04
C 19	33	1.4	192969 26	AC012359 Homo sapiens clone NHO	1.52e-04
C 20	32	1.4	198363 27	AC013545 Homo sapiens chromosom	6.60e-04
C 21	32	1.4	210296 23	HS1057B20 Homo sapiens chromosom	6.60e-04
C 22	32	1.4	247332 22	CEY18D10 Caenorhabditis elegans	6.60e-04
C 23	33	1.4	301690 28	AC010909 Homo sapiens clone RP1	1.52e-04
C 24	30	1.3	318 31	PR081266 Prionallurus bengalensis	1.14e-02
C 25	29	1.3	61613 24	AC008457 Homo sapiens chromosom	4.53e-02
C 26	30	1.3	97480 25	AC011993 Homo sapiens clone NHO	1.14e-02
C 27	31	1.3	102324 23	HS1168M15 Homo sapiens chromosom	2.79e-03
C 28	30	1.3	105525 22	AP000619 Homo sapiens chromosom	1.14e-02
C 29	30	1.3	126534 24	AC008750 Homo sapiens chromosom	1.14e-02
C 30	30	1.3	161428 24	AC009542 Homo sapiens chromosom	1.14e-02
C 31	30	1.3	165732 28	AC010794 Homo sapiens clone RP1	1.14e-02
C 32	31	1.3	166980 24	AC009756 Homo sapiens clone 45	2.79e-03
C 33	30	1.3	167886 28	AC009985 Homo sapiens clone RP1	1.14e-02
C 34	31	1.3	170953 24	AC008278 Homo sapiens clone NHO	2.79e-03
C 35	30	1.3	172200 24	AC007219 Homo sapiens chromosom	1.14e-02
C 36	31	1.3	174870 24	AC009786 Homo sapiens clone 44	2.79e-03
C 37	30	1.3	176781 26	AC015802 Homo sapiens chromosom	1.14e-02
C 38	30	1.3	180406 28	AC017004 Homo sapiens clone RP1	1.14e-02
C 39	31	1.3	182024 23	AL133258 Homo sapiens chromosom	2.79e-03
C 40	30	1.3	184062 22	AP000842 Homo sapiens chromosom	1.14e-02
C 41	30	1.3	187886 25	AC011851 Homo sapiens chromosom	1.14e-02
C 42	31	1.3	189045 28	AC010862 Homo sapiens clone RP1	2.79e-03
C 43	30	1.3	190768 22	HSJ190J20 Homo sapiens chromosom	1.14e-02
C 44	31	1.3	195874 27	AC011676 Homo sapiens clone RP1	2.79e-03
C 45	30	1.3	318663 22	HSDJ654A7 Homo sapiens chromosom	1.14e-02

ALIGNMENTS

1	RESULT	AC011389	13574 bp	DNA	HTG	06-OCT-1999
	LOCUS	Homo sapiens chromosome 5 clone C1978SKB_174D11				*** SEQUENCING IN
	DEFINITION	PROGRESS ***	1 unordered pieces.			
	ACCESSION	AC011389	1	GI:6013552		
	VERSION	AC011389.1				
	KEYWORDS	HTG; HTGS_PHASE1.				
	SOURCE	human.				
	ORGANISM	Homo sapiens				
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	TITLE	1 (bases 1 to 13574)				
	JOURNAL	DOE Joint Genome Institute.				
	REFERENCE	Sequencing of Human Chromosome 5				
	AUTHORS	Unpublished				
	TITLE	2 (bases 1 to 13574)				
	JOURNAL	DOE Joint Genome Institute.				
	REFERENCE	Direct Submission				
	AUTHORS	Submitted (06-OCT-1999)				
	TITLE	Production Sequencing Facility, DOE Joint				
	JOURNAL	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
	COMMENT	www.jgi.doe.gov.				
		* NOTE: This is a 'working draft' sequence. It currently				
		* consists of 1 contigs. The true order of the pieces				
		* is not known and their order in this sequence record is				
		* arbitrary. Gaps between the contigs are represented as				
		* runs of N, but the exact sizes of the gaps are unknown.				
		* This record will be updated with the finished sequence				
		* as soon as it is available and the accession number will				
		* be preserved.				
		* 1 13574: contig of 13574 bp in length.				
		* Location/Qualifiers				
		1. .13574				
		/organism="Homo sapiens"				
		/gb_xref="taxon:9606"				
		/chromosome="5"				
		/clone="C1978SKB_174D11"				
		3467 a 2760 c 115 g 462 t 6770 others				

BASE COUNT


```
QY 1324 TGTGAACTTGTGCTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCCACC 1383
Db 6946 DCAADADCBACBACADACCHTRANGTHWTKRANTHSDATANCBAADCBDBCADA 7005
QY 1384 TCGCTGCCCGCTCTCTTATTCATTCAGCTGTGAGCTGTGCAGAGAAG-GCAAGAACCCTTCTGA 1442
Db 7006 DCAAAADDDADDBAADBADBCDDCCABDDDDCCCBADCBBDADAADCCADAA 7065
QY 1443 AGCATTTTCGATTC-C-GACTTGGGAAG-AGGATCCCTAGCTCTCTGTTATCCCGGTG 1499
Db 7066 CAACDAAABACDCAABDDDDDDDDDBDDDBDADADCDDBDAABCAABDBDD 7125
QY 1500 -AAGTATGCTATAGCAGTATCCCGGCTTCAGACATCCCTCAGTGAATTCAGC 1558
Db 7126 ADDAADAADCCACACBBAADDDADAABDDCAABACCCBACBBBCACDDCAADA 7185
QY 1559 AAGCCGAGTTTAAAGATATAAGATCATGGCTATTCCATA-CGCACCATAGACTATAG 1617
Db 7186 DDDDBCCCCBACDDCBACAAACCBDBCCCAACCCCBADADADCBACBBACCCBBA 7245
QY 1618 GTATACGTGTGGGTGGCTTCAATCTCTGATGAATTTCTAGTAACTTTCTGACATCCA 1677
Db 7246 BABCCBCDACAACDCAACBCCBDBBACCCBDBCBDBCBDBCBDBCBDBCBDBCBDB 7305
QY 1678 TGCAGGGGAAGTGTATTTTGTGGATCTGCACCATTCGAGGATCACAAT-ATGTATATG 1736
Db 7306 DBACBBADBBACADDDCDADCBABDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBB 7365
QY 1737 ATTCCCAAGGTGGAGATCTTTTTCAGTGTGTGATGCTTGTGATGCTTGTGCAACCATGGAT 1796
Db 7366 CBAADCCBDADDBCCCBACACDABDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBB 7425
QY 1797 GCAATGTGTGTGCTCCTCCAGCTGGTGAGAGGAGGTAGAGTGTGCTGCTTTGT 1856
Db 7426 ADCBBABACACACBCCBCCBACBDBCCBDBBDBBDBBDBBDBBDBBDBBDBBDBB 7485
QY 1857 GATTACCATTAATTTGGAAGCAGCCTGAGGCTAGTTTAAATCCAAACATGCATCAACA 1916
Db 7486 ABDACABDA-ADBCBADDDCDDBADCCDDDBBDBBDBBDBBDBBDBBDBBDBBDBBDB 7544
QY 1917 TTGCGCTCAGATATGTAAAGACCAAACTTTTCTGTAGTCTTTATTAATAATTTATAT 1976
Db 7545 ABDACBDBACCBACBACBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDB 7604
QY 1977 TGTAATTTGGACCAAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTATT 2036
Db 7605 ABADADDACACACBACBACBACBACBACBACBACBACBACBACBACBACBACBACB 7664
QY 2037 ACTGAATAAATACAA-AGCAAAACAACTCAAGTTATGTATACCTTTGGATACGA-AGA- 2093
Db 7665 BCDCBDAADAAACDCDCDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 7724
QY 2094 CCATACATAATAACCAACATAACATATATACAAAGAACTTTCATTTATTTTGGGAAT 2153
Db 7725 DBBBBAADCCCBDAACACBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDB 7784
QY 2154 TTAGTGATTTCAAAAAGTAATCATATATCAAACTAGGCACACACATAGTTCCGTGATTA 2213
Db 7785 BDABADCCACABDBABDBCCCBABDBBDBBDBBDBBDBBDBBDBBDBBDBBDBBDBB 7842
QY 2214 TTTTGTATTAATTAATAATATATATCTTTAGCGCCCTATATATTCAAAATATATGTTAA 2273
Db 7843 CADDBADDBADD 7855
QY 2274 CATGTAATCCATG 2286

RESULT 3
LOCUS AC015390 16310 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC015390
VERSION AC015390.1 GI:6435945
```

```
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 16310)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10209831 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source Location/Qualifiers
1..16310
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 4280 a 3707 c 3728 g 4595 t
ORIGIN
Query Match 3.3%; Score 75; DB 26; Length 16310;
Best Local Similarity 65.6%; Pred. No. 3.22e-37;
Matches 158; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
Db 11701 AGCAATCCCTGTGTGCACCCAGCCGAATTCGCTGCTCAGGGTCAAGACCCGGATCTC 11760
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 363 AGCAAGCAGTGTGCGCCGCGGCGGCTTCTTCTCCTACTGGCAGGAGACCTGCACCA 422
Db 11761 TGCATCTGTACGACTTCTACAGTTACTGCGCAGCTTTCACCGGAATTTTCACCACTTTC 11820
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 423 CCGCGCTGTACGACTTCAACTCTCTACTGGAGGTGCACCGTGGAAACTTCTCCACCATC 482
Db 11821 CGCAGTACTTCAAGGACGACGATATACACCTATAGCTGTGGAAAGGCTTCCATCCTG 11880
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 CCGAGTACTTCAAGGAGATGCTATGTACCATGTGCGTGGGAAAGTCTTTCACCTG 542
Db 11881 GCGCTCATCTCAACACACGAGTACTATCCCTTTAGTTGGTCCGCTCCAGCATTTTCGTC 11940
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 543 GGATATCTTCTAACCATACCGATGATTCCTCCGTATAGTGGTCTTTTCCACCTTATCATC 602
Db 11941 C 11941
QY 603 C 603

RESULT 4
LOCUS AC004358 113816 bp DNA HTG 30-JUL-1999
DEFINITION Drosophila melanogaster chromosome 3 clone DS01859 (D262) map
63C5-63D3 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***
unordered pieces.
ACCESSION AC004358 AC003918 AC003919 AC003920 AC003921
VERSION AC004358.3 GI:5656737
KEYWORDS HTG: HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 113816)
AUTHORS Ceiniker,S.E., Agbayan,A., Arcaluna,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 113816)
Celniker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Humast, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.

TITLE
JOURNAL

Direct Submission
Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT

On Jul 30, 1999 this sequence version replaced gi:5597062.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. P1 library location:
20-35.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 524: contig of 524 bp in length
* 525 604: gap of unknown length
* 1220 1219: contig of 615 bp in length
* 1300 1299: gap of unknown length
* 1514 1513: contig of 214 bp in length
* 1594 1593: gap of unknown length
* 1944 1943: contig of 350 bp in length
* 2024 2023: gap of unknown length
* 2557 2556: contig of 533 bp in length
* 2637 2636: gap of unknown length
* 2928 2927: contig of 291 bp in length
* 3008 3007: gap of unknown length
* 3440 3439: contig of 432 bp in length
* 3520 3519: gap of unknown length
* 4134 4133: contig of 614 bp in length
* 4214 4832: contig of 619 bp in length
* 4833 4912: gap of unknown length
* 4913 5973: contig of 1061 bp in length
* 5974 6053: gap of unknown length
* 6054 6511: contig of 458 bp in length
* 6512 6591: gap of unknown length
* 6592 13907: contig of 7316 bp in length
* 13908 13987: gap of unknown length
* 13988 22128: contig of 8141 bp in length
* 22139 22208: gap of unknown length
* 22209 46184: contig of 23976 bp in length
* 46185 46264: gap of unknown length
* 46265 86697: contig of 40433 bp in length
* 86698 86777: gap of unknown length
* 86778 87035: contig of 258 bp in length
* 87036 87115: gap of unknown length
* 87116 87321: contig of 206 bp in length
* 87322 87401: gap of unknown length
* 87402 87878: contig of 477 bp in length
* 87879 87958: gap of unknown length
* 87959 88359: contig of 401 bp in length
* 88360 88439: gap of unknown length
* 88440 89001: contig of 562 bp in length
* 89002 89081: gap of unknown length
* 89082 89486: contig of 405 bp in length
* 89487 89566: gap of unknown length
* 89567 89925: contig of 359 bp in length
* 89926 90005: gap of unknown length
* 90006 90390: contig of 385 bp in length

90391 90470: gap of unknown length
90471 90875: contig of 405 bp in length
90876 90955: gap of unknown length
90956 91405: contig of 450 bp in length
91406 91485: gap of unknown length
91486 91901: contig of 416 bp in length
91902 91981: gap of unknown length
91982 92394: contig of 413 bp in length
92395 92474: gap of unknown length
92475 92870: contig of 396 bp in length
92871 92950: gap of unknown length
92951 93412: contig of 462 bp in length
93413 93492: gap of unknown length
93493 93915: contig of 423 bp in length
93916 93995: gap of unknown length
93996 94222: contig of 227 bp in length
94223 94302: gap of unknown length
94303 94622: contig of 320 bp in length
94623 94702: gap of unknown length
94703 94941: contig of 239 bp in length
94942 95021: gap of unknown length
95022 95412: contig of 391 bp in length
95413 95492: gap of unknown length
95493 95816: contig of 324 bp in length
95817 95896: gap of unknown length
95897 96153: contig of 257 bp in length
96154 96233: gap of unknown length
96234 96683: contig of 450 bp in length
96684 96764: gap of unknown length
96765 97249: contig of 486 bp in length
97250 97329: gap of unknown length
97330 97676: contig of 347 bp in length
97677 97756: gap of unknown length
97757 98253: contig of 497 bp in length
98254 98333: gap of unknown length
98334 98884: contig of 551 bp in length
98885 98964: gap of unknown length
98965 99282: contig of 318 bp in length
99283 99362: gap of unknown length
100020 100099: gap of unknown length
100100 100353: contig of 254 bp in length
100354 100433: gap of unknown length
100434 100993: contig of 560 bp in length
100994 101073: gap of unknown length
101074 101358: contig of 285 bp in length
101359 101438: gap of unknown length
101439 101851: contig of 413 bp in length
101852 101931: gap of unknown length
101932 102206: contig of 275 bp in length
102207 102286: gap of unknown length
102287 102567: contig of 281 bp in length
102568 102647: gap of unknown length
102648 103081: contig of 434 bp in length
103082 103161: gap of unknown length
103162 103605: contig of 444 bp in length
103606 103685: gap of unknown length
103686 104194: contig of 509 bp in length
104195 104274: gap of unknown length
104275 104672: contig of 398 bp in length
104673 104752: gap of unknown length
104753 105076: contig of 324 bp in length
105077 105156: gap of unknown length
105157 105494: contig of 338 bp in length
105495 105574: gap of unknown length
105575 105891: contig of 317 bp in length
105892 105971: gap of unknown length
105972 106391: contig of 420 bp in length
106392 106471: gap of unknown length
106472 106868: contig of 397 bp in length
106869 106948: gap of unknown length
106949 107593: contig of 645 bp in length
107594 107673: gap of unknown length

```

Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES
  source
    1..1056
      /organism="Mustela vison"
      /db_xref="taxon:9667"
      /chromosome="1"
      /map="1q"
      /note="primers: 1167F: agccctgcgtatctactctt, 1167R:
        gagcatcttcacgcgtgttgag"
    primer_bind
      98..119
        /standard_name="1167F"
        complement(300..320)
      primer_bind
        /standard_name="1167R"
BASE COUNT      211 a      221 c      210 g      225 t      189 others
ORIGIN
Query Match      1.7%; Score 40; DB 31; Length 1056;
Best Local Similarity 17.6%; Pred. No. 2.73e-09;
Matches 30; Conservative 80; Mismatches 58; Indels 2; Gaps 2;
Db 553 AYTPMGMSSCKWTTWTSGCCCRHVDVBMWKNWCKKSCCKSTGKDKMSGCAYCGKR 612
Qy 948 ACGTCCAAGCGCTTAACATCAGTGTGCCGTATGGTCCAATTCTGTGGACTTTTCAGCGGA 1007
Db 613 MRRCRYMMRMVCGSGTWARCCCDKSKSGDKSHCKSRKKYKWMDDRYHBCKSMCAWVR 672
Qy 1008 AAATCCGCCAGAGACTATTCGCTCTGTGTC-ATATTGTGATACACAGGTC-GGCCGCCT 1065
Db 673 GKWSAGTGCMAWGWRWTDSSCTSTKGARRRKGMYARRKMYGTGAANAASAT 722
Qy 1066 CTTTGAGTGCCTTGACCATCTTCAGCTGGCCACACAGCACCATTATGTCAT 1115

```

..
Note: remainder of annotations omitted.

6	AF012089	10772 bp	DNA	INV	30-JUL-1999
LOCUS	Drosophila melanogaster cysteine proteinase-1 (CP1) and phenylalanyl tRNA synthetase genes, complete cds.				
DEFINITION					
ACCESSION	AF012089				
VERSION	AF012089.1 GI:2305220				
KEYWORDS	fruit fly.				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Gray,Y.H., Tanaka,M.M. and Sved,J.A.				
AUTHORS	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion				
TITLE	Genetics 144 (4), 1601-1610 (1996)				
JOURNAL	79132596				
MEDLINE	2 (bases 1 to 10772)				
REFERENCE	Gray,Y.H., Sved,J.A., Preston,C.R. and Engels,W.R.				
AUTHORS	Structure and associated mutational effects of the cysteine proteinase (CP1) gene of Drosophila melanogaster				
TITLE	Insect Mol. Biol. 7 (3), 291-293 (1998)				
JOURNAL	98324856				
MEDLINE					

```

AUTHORS      Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.
TITLE        Direct Submission
JOURNAL      Submitted (30-JUN-1997) School of Biological Sciences, University
              of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES     Location/Qualifiers
              1. .10772
                 /organism="Drosophila melanogaster"
                 /db_xref="taxon:7227"
mRNA         join(872..1000,2310..2426,6476..6690,6751..7707)
              /gene="Cp1"
              /product="cysteine protease"
gene         872..7707
              /gene="Cp1"
exon        872..1000
              /gene="Cp1"

```


[illegible]

```
/db_xref="GI:5656618"
/translation="MLTLRVQGARHWLKSTRCLASAAAPAKSPSPQPLEVSGSTYA
TDGNTVTPKILSPGANKHLQDHPILSIIRQIVNYFYGAYRNQRGNPLFSVYDQMN
PVYVQOQFNLLIPADHVSRQSDCYIINQOHLRAHTTAHQVELISGGLDNFLVVG
EYVRROIDSHTYVFOADAVRLVTKDLFPNGLFELFEATWSGTGLADPKLILPH
SSYTKPNSPATRAKVLMEHMKHVLGLTKDLFGPRIKYRWVDTYFPFQPSWELEI
YFDNWLIEGLGIMRHEILQRSVHQSIGAFGVGLERLAWLFDLPIDIRLFSNDS
GFLSQSEKLDHLLPKYKPIKSHYPOCTNDLSFPLQDIEVDAGFSPNDFYDLVRSVAG
DMVEQISLVDFKHPRKTGKSSVCVRIYRHMERTLQAEVNEIHQIASASVDSENVQ
IR"
BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
ORIGIN

Query Match 1.7%; Score 38; DB 30; Length 10772;
Best Local Similarity 16.2%; Pred. No. 6.90e-08;
Matches 22; Conservative 65; Mismatches 49; Indels 0; Gaps 0;

Db 1691 AKRWYNAWTTMKWMMTWKWKMMKYRTWMMKMTWTSTRTTTSAMWMMYTWSTWK 1750
|::: |:| ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 2076 ATGACATAACTGTGAGTTTGTGCTTTGCTTATTTATTCAGTAAATAAGCGCTAACTGTTT 2017
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1751 YWYAWMMKMTWTWARMANASWARKWKTSAAAAYSARWKWKWYARWAMKKTWMAA 1810
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 2016 AAAAGAGGGAATAAAAAAAACTGCTCAATTACCAATTATAAATTTAATAAAGA 1957
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1811 WKRWKAAMTWRRWY 1826
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 1956 CTAACGAGAAAGGTTT 1941
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 8
LOCUS MYU87256 1056 bp DNA MAM 05-JAN-1999
DEFINITION Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION U87256
VERSION U87256.1 GI:4099442
KEYWORDS
SOURCE American mink.
ORGANISM Mustela vison
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE 1 (bases 1 to 1056)
AUTHORS Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES
source
1..1056
/organism="Mustela vison"
/db_xref="taxon:9667"
/chromosome="1"
/map="1q"
/notes="primers: 1167F: agcccttgcatatctactctt, 1167R:
gagatattaccgctgtgag"
98..119
/standard_name="1167F"
complement(300..320)
/standard_name="1167R"
BASE COUNT 211 a 221 c 210 g 225 t 189 others
ORIGIN

Query Match 1.6%; Score 37; DB 31; Length 1056;
Best Local Similarity 15.5%; Pred. No. 3.36e-07;
Matches 31; Conservative 90; Mismatches 78; Indels 1; Gaps 1;

Db 541 CATCKGGSGGTAYTWMGMSKRCWTWTSGCCWSCRHDVBMWSKWCGRKSCCKSTGK 600
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 928 CATCCAGGGTGTAGGCCACAGGGGTAGCCATCAGGACCTCGGATCGGGGCCAG 869
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 601 DKMSGCYGKMRRCRSYMMRWCMVSGTWARCCCDKSGSGDKSHCKSKRYKWMDR 660
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 868 GGTGATGTTCTCCCAAGGATACAACTTCTGAATTCCTTGGGTATCTGAAGGGGATGTG 809
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 661 YHBCKSWCMYRGKWSAGTCGMARGWRTWDSSTCTGKARRRKNYAR-RKMYGTGAWA 719
```

```
Cp 808 TGCGTTATGATCCCAACGCCAGGAAGAAGACTGGCTGAGTTTTCATCTTTCCAA 749
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 720 SATAYSTGATGTTBWRVTSWK 739
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Cp 748 CAACGTATGGCTTGCTCAG 729
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 9
LOCUS AC016207 56722 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens clone Rp11-25H20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016207
VERSION AC016207.1 GI:6466587
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone Rp11-25H20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Harford,A., Klein,J.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Kohn,L.,
Lewczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4574
Center clone name: 25_H_20
-----
* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 778: contig of 778 bp in length
* * gap of unknown length
* 779 1539: contig of 761 bp in length
* * gap of unknown length
* 1540 2321: contig of 782 bp in length
* * gap of unknown length
* 2322 3075: contig of 754 bp in length
* * gap of unknown length
* 3076 3885: contig of 810 bp in length
* * gap of unknown length
```


Db 35224 AATGTAATTTTTT 35236
 QY 2283 CATGTCCTTTTT 2295

```

RESULT 10
LOCUS AC016207 56722 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens clone RP11-25H20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016207
VERSION AC016207.1 GI:6466587
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-25H20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukigalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIFBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4574
Center clone name: 25_H_20
-----
* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 778: contig of 778 bp in length
* gap of unknown length
* 779 1539: contig of 761 bp in length
* gap of unknown length
* 1540 2321: contig of 782 bp in length
* gap of unknown length
* 2322 3075: contig of 754 bp in length
* gap of unknown length
* 3076 3885: contig of 810 bp in length
* gap of unknown length
* 3886 4699: contig of 814 bp in length
* gap of unknown length

```

```

* 4700 5482: contig of 783 bp in length
* gap of unknown length
* 5483 6252: contig of 770 bp in length
* gap of unknown length
* 6253 7020: contig of 768 bp in length
* gap of unknown length
* 7021 7844: contig of 824 bp in length
* gap of unknown length
* 7845 8678: contig of 834 bp in length
* gap of unknown length
* 8679 9498: contig of 820 bp in length
* gap of unknown length
* 9499 10298: contig of 800 bp in length
* gap of unknown length
* 10299 11082: contig of 784 bp in length
* gap of unknown length
* 11083 11865: contig of 783 bp in length
* gap of unknown length
* 11866 12638: contig of 773 bp in length
* gap of unknown length
* 12639 13435: contig of 797 bp in length
* gap of unknown length
* 13436 14219: contig of 784 bp in length
* gap of unknown length
* 14220 15013: contig of 794 bp in length
* gap of unknown length
* 15014 15763: contig of 750 bp in length
* gap of unknown length
* 15764 16528: contig of 765 bp in length
* gap of unknown length
* 16529 17296: contig of 768 bp in length
* gap of unknown length
* 17297 18069: contig of 773 bp in length
* gap of unknown length
* 18070 18855: contig of 786 bp in length
* gap of unknown length
* 18856 19641: contig of 786 bp in length
* gap of unknown length
* 19642 20497: contig of 856 bp in length
* gap of unknown length
* 20498 21280: contig of 783 bp in length
* gap of unknown length
* 21281 22064: contig of 784 bp in length
* gap of unknown length
* 22065 22848: contig of 784 bp in length
* gap of unknown length
* 22849 23639: contig of 791 bp in length
* gap of unknown length
* 23640 24419: contig of 780 bp in length
* gap of unknown length
* 24420 25193: contig of 774 bp in length
* gap of unknown length
* 25194 25972: contig of 779 bp in length
* gap of unknown length
* 25973 26744: contig of 772 bp in length
* gap of unknown length
* 26745 27520: contig of 776 bp in length
* gap of unknown length
* 27521 28261: contig of 741 bp in length
* gap of unknown length
* 28262 29052: contig of 791 bp in length
* gap of unknown length
* 29053 29842: contig of 790 bp in length
* gap of unknown length
* 29843 30623: contig of 781 bp in length
* gap of unknown length
* 30624 31411: contig of 788 bp in length
* gap of unknown length
* 31412 32225: contig of 814 bp in length
* gap of unknown length
* 32226 33057: contig of 832 bp in length
* gap of unknown length
* 33058 33839: contig of 782 bp in length

```

||||| |||

Cp 2236 ATATTATTAATT 2224

RESULT 11

LOCUS AC016220 59006 bp DNA HTG 23-NOV-1999

DEFINITION Homo sapiens clone RP11-26H10, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC016220

VERSION AC016220.1 GI:6466574

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 59006)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens, clone RP11-26H10

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 59006)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Farreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galanad,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatis,A., Klein,J.,
Lewandzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

TITLE Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

FEATURES	source	Location/Qualifiers	1. .56722	/organism="Homo sapiens"	...	Note: remainder of annotations omitted.
4935	*	50420:	contig of 786 bp in length	gap of unknown length	1.5%	Score 34; DB 27; Length 56722;
50421	*	51195:	contig of 775 bp in length	gap of unknown length	76.7%	Pred. No. 3.42e-05;
51196	*	51980:	contig of 785 bp in length	gap of unknown length	0;	Mismatches 16; Indels 1; Gaps 1;
51981	*	52798:	contig of 808 bp in length	gap of unknown length		
52789	*	53601:	contig of 813 bp in length	gap of unknown length		
53602	*	54382:	contig of 781 bp in length	gap of unknown length		
54383	*	55162:	contig of 780 bp in length	gap of unknown length		
55163	*	55944:	contig of 782 bp in length	gap of unknown length		
55945	*	56722:	contig of 778 bp in length.			
Db	55644	AAAAAATACATTGATTACATGTTGAATCATATTTTGGAAATATTAACGTTTAATGAAT	55703			
Cp	2295	AAAANGAATCATGGATTACATGTTTAACATATATTTTGAATATAT-AGGCTCATAGAT	2237			
Db	55704	ATATTATTGATTT	55716			

```
* 5391 6168: contig of 778 bp in length
*      gap of unknown length
* 6169 6945: contig of 777 bp in length
*      gap of unknown length
* 6946 7725: contig of 780 bp in length
*      gap of unknown length
* 7726 8512: contig of 787 bp in length
*      gap of unknown length
* 8513 9286: contig of 774 bp in length
*      gap of unknown length
* 9287 10068: contig of 782 bp in length
*      gap of unknown length
* 10069 10840: contig of 772 bp in length
*      gap of unknown length
* 10841 11624: contig of 784 bp in length
*      gap of unknown length
* 11625 12408: contig of 784 bp in length
*      gap of unknown length
* 12409 13186: contig of 778 bp in length
*      gap of unknown length
* 13187 13982: contig of 796 bp in length
*      gap of unknown length
* 13983 14739: contig of 757 bp in length
*      gap of unknown length
* 14740 15503: contig of 764 bp in length
*      gap of unknown length
* 15504 16285: contig of 782 bp in length
*      gap of unknown length
* 16286 17071: contig of 786 bp in length
*      gap of unknown length
* 17072 17836: contig of 765 bp in length
*      gap of unknown length
* 17837 18613: contig of 777 bp in length
*      gap of unknown length
* 18614 19392: contig of 779 bp in length
*      gap of unknown length
* 19393 20186: contig of 794 bp in length
*      gap of unknown length
* 20187 20973: contig of 787 bp in length
*      gap of unknown length
* 20974 21761: contig of 788 bp in length
*      gap of unknown length
* 21762 22531: contig of 770 bp in length
*      gap of unknown length
* 22532 23328: contig of 797 bp in length
*      gap of unknown length
* 23329 24093: contig of 765 bp in length
*      gap of unknown length
* 24094 24848: contig of 755 bp in length
*      gap of unknown length
* 24849 25631: contig of 783 bp in length
*      gap of unknown length
* 25632 26419: contig of 788 bp in length
*      gap of unknown length
* 26420 27211: contig of 792 bp in length
*      gap of unknown length
* 27212 27985: contig of 774 bp in length
*      gap of unknown length
* 27986 28706: contig of 721 bp in length
*      gap of unknown length
* 28707 29461: contig of 755 bp in length
*      gap of unknown length
* 29462 30242: contig of 781 bp in length
*      gap of unknown length
* 30243 31021: contig of 779 bp in length
*      gap of unknown length
* 31022 31784: contig of 763 bp in length
*      gap of unknown length
* 31785 32568: contig of 784 bp in length
*      gap of unknown length
* 32569 33354: contig of 786 bp in length
*      gap of unknown length
* 33355 34124: contig of 770 bp in length
```

```
* 34125 34907: gap of unknown length
*      contig of 783 bp in length
* 34908 35673: contig of 766 bp in length
*      gap of unknown length
* 35674 36429: contig of 756 bp in length
*      gap of unknown length
* 36430 37192: contig of 763 bp in length
*      gap of unknown length
* 37193 37952: contig of 760 bp in length
*      gap of unknown length
* 37953 38730: contig of 778 bp in length
*      gap of unknown length
* 38731 39538: contig of 808 bp in length
*      gap of unknown length
* 39539 40316: contig of 778 bp in length
*      gap of unknown length
* 40317 41090: contig of 774 bp in length
*      gap of unknown length
* 41091 41870: contig of 780 bp in length
*      gap of unknown length
* 41871 42662: contig of 792 bp in length
*      gap of unknown length
* 42663 43426: contig of 764 bp in length
*      gap of unknown length
* 43427 44186: contig of 760 bp in length
*      gap of unknown length
* 44187 44957: contig of 771 bp in length
*      gap of unknown length
* 44958 45744: contig of 787 bp in length
*      gap of unknown length
* 45745 46518: contig of 774 bp in length
*      gap of unknown length
* 46519 47306: contig of 788 bp in length
*      gap of unknown length
* 47307 48098: contig of 792 bp in length
*      gap of unknown length
* 48099 48887: contig of 789 bp in length
*      gap of unknown length
* 48888 49664: contig of 777 bp in length
*      gap of unknown length
* 49665 50451: contig of 787 bp in length
*      gap of unknown length
* 50452 51233: contig of 782 bp in length
*      gap of unknown length
* 51234 51982: contig of 749 bp in length
*      gap of unknown length
* 51983 52771: contig of 789 bp in length
*      gap of unknown length
* 52772 53527: contig of 756 bp in length
*      gap of unknown length
* 53528 54314: contig of 787 bp in length
*      gap of unknown length
* 54315 55085: contig of 771 bp in length
*      gap of unknown length
* 55086 55893: contig of 808 bp in length
*      gap of unknown length
* 55894 56677: contig of 784 bp in length
*      gap of unknown length
```

... Note: remainder of annotations omitted.

Query Match 1.5%; Score 34; DB 27; Length 59006;
Best Local Similarity 80.0%; Pred. NO. 3.42e-05;
Matches 52; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Db 15973 ATGTGATTACACATTACATAATATTTTGGAGATAATTTGCTAAATAATATATTATTAA 16032

Cp 2285 ATGGATTACATCTTAACATAATATTTTGAATATATAGGCTCA-TAAGATATATTATTAA 2227

Db 16033 ATTAT 16037

Cp 2226 ATTAT 2222

```

RESULT 12
LOCUS AC016439 67656 bp DNA HTG 26-NOV-1999
DEFINITION Homo sapiens clone RP11-30C21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016439
VERSION AC016439.1 GI:6468800
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 67656)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-30C21
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 67656)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4861
Center clone name: 30_C_21
-----
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 741: contig of 741 bp in length
gap of unknown length
742 1519: contig of 778 bp in length
gap of unknown length
1520 2316: contig of 797 bp in length
gap of unknown length
2317 3103: contig of 787 bp in length
gap of unknown length
3104 3870: contig of 767 bp in length
gap of unknown length
3871 4640: contig of 770 bp in length
gap of unknown length
4641 5433: contig of 793 bp in length
gap of unknown length
5434 6223: contig of 790 bp in length
gap of unknown length
6224 6976: contig of 753 bp in length
gap of unknown length
6977 7737: contig of 761 bp in length
gap of unknown length
7738 8503: contig of 766 bp in length
gap of unknown length
8504 9287: contig of 784 bp in length
gap of unknown length
9288 10066: contig of 779 bp in length
gap of unknown length
10067 10852: contig of 786 bp in length
gap of unknown length
10853 11621: contig of 769 bp in length
gap of unknown length
11622 12411: contig of 790 bp in length
gap of unknown length
12412 13191: contig of 780 bp in length
gap of unknown length
13192 13983: contig of 792 bp in length
gap of unknown length
13984 14772: contig of 789 bp in length
gap of unknown length
14773 15538: contig of 766 bp in length
gap of unknown length
15539 16299: contig of 761 bp in length
gap of unknown length
16300 17091: contig of 792 bp in length
gap of unknown length
17092 17829: contig of 738 bp in length
gap of unknown length
17830 18630: contig of 801 bp in length
gap of unknown length
18631 19415: contig of 785 bp in length
gap of unknown length
19416 20205: contig of 790 bp in length
gap of unknown length
20206 20965: contig of 760 bp in length
gap of unknown length
20966 21730: contig of 765 bp in length
gap of unknown length
21731 22511: contig of 781 bp in length
gap of unknown length
22512 23290: contig of 779 bp in length
gap of unknown length
23291 24068: contig of 778 bp in length
gap of unknown length
24069 24841: contig of 773 bp in length
gap of unknown length
24842 25616: contig of 775 bp in length
gap of unknown length
25617 26401: contig of 785 bp in length
gap of unknown length
26402 27160: contig of 759 bp in length
gap of unknown length
27161 27919: contig of 759 bp in length
gap of unknown length
27920 28716: contig of 797 bp in length
gap of unknown length
28717 29511: contig of 795 bp in length
gap of unknown length
29512 30270: contig of 759 bp in length
gap of unknown length
30271 31050: contig of 780 bp in length
gap of unknown length
31051 31819: contig of 769 bp in length
gap of unknown length
31820 32597: contig of 778 bp in length
gap of unknown length
32598 33379: contig of 782 bp in length
gap of unknown length
33380 34168: contig of 789 bp in length
gap of unknown length
34169 34930: contig of 762 bp in length
gap of unknown length

```

*	34931	35724:	gap of 794 bp in length
*			contig of unknown length
*	35725	36518:	gap of 794 bp in length
*			contig of unknown length
*	36519	37298:	gap of 780 bp in length
*			contig of unknown length
*	37299	38064:	gap of 766 bp in length
*			contig of unknown length
*	38065	38842:	gap of 778 bp in length
*			contig of unknown length
*	38843	39626:	gap of 784 bp in length
*			contig of unknown length
*	39627	40411:	gap of 785 bp in length
*			contig of unknown length
*	40412	41198:	gap of 787 bp in length
*			contig of unknown length
*	41199	41979:	gap of 781 bp in length
*			contig of unknown length
*	41980	42766:	gap of 787 bp in length
*			contig of unknown length
*	42767	43528:	gap of 762 bp in length
*			contig of unknown length
*	43529	44321:	gap of 793 bp in length
*			contig of unknown length
*	44322	45110:	gap of 789 bp in length
*			contig of unknown length
*	45111	45872:	gap of 762 bp in length
*			contig of unknown length
*	45873	46648:	gap of 776 bp in length
*			contig of unknown length
*	46649	47437:	gap of 789 bp in length
*			contig of unknown length
*	47438	48206:	gap of 769 bp in length
*			contig of unknown length
*	48207	48976:	gap of 770 bp in length
*			contig of unknown length
*	48977	49764:	gap of 788 bp in length
*			contig of unknown length
*	49765	50543:	gap of 779 bp in length
*			contig of unknown length
*	50544	51338:	gap of 795 bp in length
*			contig of unknown length
*	51339	52102:	gap of 764 bp in length
*			contig of unknown length
*	52103	52893:	gap of 791 bp in length
*			contig of unknown length
*	52894	53692:	gap of 799 bp in length
*			contig of unknown length
*	53693	54475:	gap of 783 bp in length
*			contig of unknown length
*	54476	55252:	gap of 777 bp in length
*			contig of unknown length
*	55253	56011:	gap of 759 bp in length
*			contig of unknown length
*	56012	56784:	gap of 773 bp in length
*			contig of unknown length

...
Note: remainder of annotations omitted.

Query Match 1.5%; Score 35; DB 27; Length 67656;
Best Local Similarity 80.3%; Pred. No. 7.50e-06;
Matches 53; Conservative 0; Mismatches 12; Indels 1

Accession	Sequence	Length
Db 48059	TAAATTTTAATACTATATTTTATTCAACCCCAATATATCCAAAATACTAAGTCAACATGTA	48118
Oy 2221	TATAATTTTAATAATATATCTCTATG-AGCCCATATATATTCAAAATATATTATGTTAACATGTA	2279

Db 48119 ATCAAT 48124

Qv 2280 ATCCAT 2285

RESULT 13
LOCUS HSA175J10 176224 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens chromosome 10 clone RP11-175J10, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AL121927
VERSION AL121927.2 GI:6066944
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 176224)
REFERENCE Plumb,B.
AUTHORS Direct Submission
TITLE Submitted (09-NOV-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
JOURNAL humquest@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Oct 19, 1999 this sequence version replaced gi:6065911.
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: bA175J10 Contig_ID: 00007 acc=AL121927 Length: 1925 bp Unfinished: bA175J10 Contig_ID: 00081 acc=AL121927 Length: 2568 bp Unfinished: bA175J10 Contig_ID: 00094 acc=AL121927 Length: 1018 bp Unfinished: bA175J10 Contig_ID: 00112 acc=AL121927 Length: 4164 bp Unfinished: bA175J10 Contig_ID: 00135 acc=AL121927 Length: 1776 bp Unfinished: bA175J10 Contig_ID: 00137 acc=AL121927 Length: 2467 bp Unfinished: bA175J10 Contig_ID: 00138 acc=AL121927 Length: 1787 bp Unfinished: bA175J10 Contig_ID: 00191 acc=AL121927 Length: 2684 bp Unfinished: bA175J10 Contig_ID: 00220 acc=AL121927 Length: 1288 bp Unfinished: bA175J10 Contig_ID: 00227 acc=AL121927 Length: 1822 bp Unfinished: bA175J10 Contig_ID: 00241 acc=AL121927 Length: 3671 bp Unfinished: bA175J10 Contig_ID: 00273 acc=AL121927 Length: 3418 bp Unfinished: bA175J10 Contig_ID: 00319 acc=AL121927 Length: 1384 bp Unfinished: bA175J10 Contig_ID: 00326 acc=AL121927 Length: 4574 bp Unfinished: bA175J10 Contig_ID: 00337 acc=AL121927 Length: 1234 bp Unfinished: bA175J10 Contig_ID: 00367 acc=AL121927 Length: 1063 bp Unfinished: bA175J10 Contig_ID: 00381 acc=AL121927 Length: 1269 bp Unfinished: bA175J10 Contig_ID: 00393 acc=AL121927 Length: 1294 bp Unfinished: bA175J10 Contig_ID: 00459 acc=AL121927 Length: 1006 bp Unfinished: bA175J10 Contig_ID: 00463 acc=AL121927 Length: 5421 bp Unfinished: bA175J10 Contig_ID: 00488 acc=AL121927 Length: 3377 bp Unfinished: bA175J10 Contig_ID: 00497 acc=AL121927 Length: 2272 bp Unfinished: bA175J10 Contig_ID: 00513 acc=AL121927 Length: 1006 bp Unfinished: bA175J10 Contig_ID: 00530 acc=AL121927 Length: 3061 bp Unfinished: bA175J10 Contig_ID: 00539 acc=AL121927 Length: 5632 bp Unfinished: bA175J10 Contig_ID: 00581 acc=AL121927 Length: 18128 bp Unfinished: bA175J10 Contig_ID: 00584 acc=AL121927 Length: 5140 bp Unfinished: bA175J10 Contig_ID: 00598 acc=AL121927 Length: 1119 bp Unfinished: bA175J10 Contig_ID: 00639 acc=AL121927 Length: 2580 bp Unfinished: bA175J10 Contig_ID: 00673 acc=AL121927 Length: 2186 bp Unfinished: bA175J10 Contig_ID: 00777 acc=AL121927 Length: 4918 bp Unfinished: bA175J10 Contig_ID: 00822 acc=AL121927 Length: 1016 bp Unfinished: bA175J10 Contig_ID: 00853 acc=AL121927 Length: 1110 bp Unfinished: bA175J10 Contig_ID: 00865 acc=AL121927 Length: 2962 bp Unfinished: bA175J10 Contig_ID: 00869 acc=AL121927 Length: 3316 bp Unfinished: bA175J10 Contig_ID: 00908 acc=AL121927 Length: 1088 bp Unfinished: bA175J10 Contig_ID: 00914 acc=AL121927 Length: 1183 bp Unfinished: bA175J10 Contig_ID: 00930 acc=AL121927 Length: 2830 bp Unfinished: bA175J10 Contig_ID: 00955 acc=AL121927 Length: 5613 bp Unfinished: bA175J10 Contig_ID: 00978 acc=AL121927 Length: 1315 bp Unfinished: bA175J10 Contig_ID: 01005 acc=AL121927 Length: 1455 bp Unfinished: bA175J10 Contig_ID: 01007 acc=AL121927 Length: 1305 bp Unfinished: bA175J10 Contig_ID: 01024 acc=AL121927 Length:

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boquslavsky, L., Boukhgalter, B., Brown, A., Castie, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeBellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

TITLE
JOURNAL
COMMENT

Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Project Information
Center project name: L625
Center clone name: 54_A_1

* NOTE: This record contains 234 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 775: contig of 775 bp in length
* gap of unknown length
* 776 contig of 758 bp in length
* gap of unknown length
* 1534 2303: contig of 770 bp in length
* gap of unknown length
* 2304 3054: contig of 751 bp in length
* gap of unknown length
* 3055 3833: contig of 779 bp in length
* gap of unknown length
* 3834 4596: contig of 763 bp in length
* gap of unknown length
* 4597 5342: contig of 746 bp in length
* gap of unknown length
* 5343 6084: contig of 742 bp in length
* gap of unknown length
* 6085 6842: contig of 758 bp in length
* gap of unknown length
* 6843 7590: contig of 748 bp in length
* gap of unknown length
* 7591 8347: contig of 757 bp in length
* gap of unknown length
* 8348 9100: contig of 753 bp in length
* gap of unknown length
* 9101 9878: contig of 778 bp in length
* gap of unknown length
* 9879 10638: contig of 760 bp in length
* gap of unknown length
* 10639 11405: contig of 767 bp in length
* gap of unknown length
* 11406 12216: contig of 811 bp in length
* gap of unknown length
* 12217 12986: contig of 770 bp in length
* gap of unknown length
* 12987 13746: contig of 760 bp in length
* gap of unknown length
* 13747 14520: contig of 774 bp in length
* gap of unknown length
* 14521 15256: contig of 736 bp in length
* gap of unknown length
* 15257 16019: contig of 763 bp in length
* gap of unknown length
* 16020 16781: contig of 762 bp in length
* gap of unknown length
* 16782 17512: contig of 731 bp in length

* 17513 18283: contig of 771 bp in length
* gap of unknown length
* 18284 19039: contig of 756 bp in length
* gap of unknown length
* 19040 19814: contig of 775 bp in length
* gap of unknown length
* 19815 20576: contig of 762 bp in length
* gap of unknown length
* 20577 21234: contig of 658 bp in length
* gap of unknown length
* 21235 21992: contig of 758 bp in length
* gap of unknown length
* 21993 22737: contig of 745 bp in length
* gap of unknown length
* 22738 23510: contig of 773 bp in length
* gap of unknown length
* 23511 24255: contig of 745 bp in length
* gap of unknown length
* 24256 25060: contig of 805 bp in length
* gap of unknown length
* 25061 25822: contig of 762 bp in length
* gap of unknown length
* 25823 26682: contig of 860 bp in length
* gap of unknown length
* 26683 27437: contig of 755 bp in length
* gap of unknown length
* 27438 28187: contig of 750 bp in length
* gap of unknown length
* 28188 28936: contig of 749 bp in length
* gap of unknown length
* 28937 29707: contig of 771 bp in length
* gap of unknown length
* 29708 30473: contig of 766 bp in length
* gap of unknown length
* 30474 31233: contig of 760 bp in length
* gap of unknown length
* 31234 31980: contig of 747 bp in length
* gap of unknown length
* 31981 32748: contig of 768 bp in length
* gap of unknown length
* 32749 33519: contig of 771 bp in length
* gap of unknown length
* 33520 34282: contig of 763 bp in length
* gap of unknown length
* 34283 35049: contig of 767 bp in length
* gap of unknown length
* 35050 35822: contig of 773 bp in length
* gap of unknown length
* 35823 36597: contig of 775 bp in length
* gap of unknown length
* 36598 37346: contig of 749 bp in length
* gap of unknown length
* 37347 38099: contig of 753 bp in length
* gap of unknown length
* 38100 38799: contig of 700 bp in length
* gap of unknown length
* 38800 39542: contig of 743 bp in length
* gap of unknown length
* 39543 40306: contig of 764 bp in length
* gap of unknown length
* 40307 41058: contig of 752 bp in length
* gap of unknown length
* 41059 41821: contig of 763 bp in length
* gap of unknown length
* 41822 42562: contig of 741 bp in length
* gap of unknown length
* 42563 43302: contig of 740 bp in length
* gap of unknown length
* 43303 44064: contig of 762 bp in length
* gap of unknown length
* 44065 44857: contig of 793 bp in length
* gap of unknown length

```
* 44858 45617: contig of 760 bp in length
* gap of unknown length
* 45618 46385: contig of 768 bp in length
* gap of unknown length
* 46386 47154: contig of 769 bp in length
* gap of unknown length
* 47155 47915: contig of 761 bp in length
* gap of unknown length
* 47916 48659: contig of 744 bp in length
* gap of unknown length
* 48660 49442: contig of 783 bp in length
* gap of unknown length
* 49443 50204: contig of 762 bp in length
* gap of unknown length
* 50205 50981: contig of 777 bp in length
* gap of unknown length
* 50982 51746: contig of 765 bp in length
* gap of unknown length
* 51747 52515: contig of 769 bp in length
* gap of unknown length
* 52516 53383: contig of 868 bp in length
* gap of unknown length
* 53384 54171: contig of 788 bp in length
* gap of unknown length
* 54172 54939: contig of 768 bp in length
* gap of unknown length
* 54940 55702: contig of 763 bp in length
* gap of unknown length
* 55703 56477: contig of 775 bp in length
* gap of unknown length
* 56478 57239: contig of 762 bp in length
* gap of unknown length
```

....
Note: remainder of annotations omitted.

```
Query Match      1.5%; Score 34; DB 26; Length 192798;
Best Local Similarity 76.6%; Pred. No. 3.42e-05;
Matches 49; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 124736 TTTAATAATATGTTGCATGACGCAAAATATACAGAAATATTATTTCAACATGTGATCAAT 124795
|||||
Qy 2226 TTTAATAATATATCTTATGAGCCCTATATATTCAAAATATTATGTTAACAATGTAATCCAT 2285
|||||

Db 124796 ATTT 124799
|||
Qy 2286 GTTT 2289
```

Search completed: Thu Apr 20 16:13:59 2000
Job time : 11804 secs.

W A R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Apr 23 09:36:33 2000; MasPar time 3972.39 Seconds
1298.715 Million cell updates/sec

Tabular output not generated:

Title: >US-09-249-003-1
Description: (1-2297) from US09249003.seq
Perfect Score: 2297
N.A. Sequence: 1 CGGCTGTCTTGGCAGTCTT.....TAATCATGTTCTTTTCC 2297
Comp: GCCGACACACCGTCAGAA.....ATTAGTACAAAGAAAGG

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 623196 seqs, 1122988046 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: genbank2-115
1:gb_pat 2:gb_ph 3:gb_pl1 4:gb_pl2 5:gb_pl3 6:gb_pr1
7:gb_pr2 8:gb_pr3 9:gb_pr4 10:gb_ro 11:gb_sts 12:gb_sy
13:gb_un 14:gb_vi

Statistics: Mean 11.531; Variance 5.769; scale 1.999

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2297	100.0	2297	6 HUMIDSX	Human iduronate 2-sulf	0.00e+00
2	2297	100.0	2297	1 I92327	Sequence 1 from patent	0.00e+00
3	2297	100.0	2297	1 AR025358	Sequence 1 from patent	0.00e+00
4	1113	48.5	1314	7 HUMIDSALT	Homo sapiens iduronate	0.00e+00
5	1096	47.7	1750	10 MUSIDS	Mus musculus iduronate	0.00e+00
6	996	43.4	1120	6 HUMIDSX	Homo sapiens iduronate	0.00e+00
7	996	43.4	4428	1 AR025360	Sequence 6 from patent	0.00e+00
8	996	43.4	4428	1 I92329	Sequence 6 from patent	0.00e+00
9	995	43.3	36845	6 HUMIDS	Homo sapiens iduronate	0.00e+00
c 10	995	43.3	326663	9 AF011889	Human Xq28 cosmid U12	0.00e+00
11	437	19.0	1098	8 AF050145	Homo sapiens iduronate	0.00e+00
12	434	18.9	1831	10 MMRNA15	M.musculus mRNA for id	0.00e+00
13	390	17.0	400	11 G13552	human STS SHGC-11149.	1.49e-290
14	340	14.8	110079	10 AC002315	Mouse BAC-146N21 Chrom	1.86e-248
c 15	231	10.1	233	1 A78222	Sequence 929 from Pate	1.60e-157
c 16	231	10.1	233	1 A75243	Sequence 929 from Pate	1.60e-157
17	229	10.0	536	6 HUMIDSX	Homo sapiens iduronate	7.22e-156
18	205	8.9	322	6 HUMISS05	Homo sapiens iduronate	4.33e-136

19 205 8.9 490 6 HUMIDSX 2297 bp mRNA PRI 07-MAR-1995
c 20 186 8.1 297 7 HSU66055 Human clone PI5a-1 IDS 4.33e-136
21 186 8.1 425 7 HSU66054 Human clone PI3b-2 IDS 1.59e-120
c 22 187 8.1 487 7 HSU66053 Human clone PI3b-1 IDS 2.43e-121
23 182 7.9 298 6 HUMISS03 Homo sapiens iduronate 2.94e-117
24 182 7.9 388 6 HUMIDSX3 Homo sapiens iduronate 2.94e-117
c 25 182 7.9 627 8 AF050146 Homo sapiens clone 488 2.94e-117
26 182 7.9 750 7 HUMIDS02 Homo sapiens iduronate 2.94e-117
27 182 7.9 326663 9 AF011889 Human Xq28 cosmid U12 2.94e-117
28 178 7.7 290 6 HUMISS06 Homo sapiens iduronate 5.35e-114
29 176 7.7 294 6 HUMISS08 Homo sapiens iduronate 2.27e-112
30 178 7.7 379 6 HUMIDSX6 Homo sapiens iduronate 5.35e-114
31 176 7.7 404 6 HUMIDSX8 Homo sapiens iduronate 2.27e-112
32 176 7.7 2331 8 HUMIDS71 Human (clone pBI7-8) i 2.27e-112
33 139 6.1 257 6 HUMISS02 Homo sapiens iduronate 1.68e-82
34 139 6.1 426 6 HUMIDSX2 Homo sapiens iduronate 1.68e-82
35 132 5.7 248 6 HUMISS07 Homo sapiens iduronate 6.46e-77
36 132 5.7 319 6 HUMIDSX7 Homo sapiens iduronate 6.46e-77
37 105 4.6 162 6 HUMISS01 Homo sapiens iduronate 1.17e-55
38 91 4.0 366 6 HUMIDSX4 Homo sapiens iduronate 7.27e-45
39 90 3.9 169 6 HUMISS04 Homo sapiens iduronate 4.21e-44
40 65 2.8 326 7 HUMIDS01 Homo sapiens iduronate 1.72e-25
c 42 47 2.0 215 1 I28278 Sequence 14 from patent 2.46e-23
43 45 2.0 74371 8 AC005369 Homo sapiens chromosom 1.31e-11
c 44 40 1.7 131389 9 AC005728 Homo sapiens Chromosom 2.23e-08
c 45 36 1.6 965 1 AR024229 Sequence 22 from patent 6.74e-06

ALIGNMENTS

RESULT 1 HUMIDSX 2297 bp mRNA PRI 07-MAR-1995
LOCUS Human iduronate 2-sulfatase mRNA, complete cds.
DEFINITION M58342 M38371
ACCESSION M58342.1 GI:184561
VERSION iduronate-2-sulfatase.
KEYWORDS Human, cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2297)
AUTHORS Willson,P.J., Morris,C.P., Anson,D.S., Occhiodoro,T., Bielicki,J.,
Clements,P.R. and Hopwood,J.J.
TITLE Hunter syndrome: Isolation of an iduronate-2-sulfatase cDNA clone
and analysis of patient DNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (21), 8531-8535 (1990)
MEDLINE 91046030
FEATURES Location/Qualifiers
source 1. .2297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="lamda c2s23"
/cell_type="endothelial"
/map="Xq27.3-q28"
125. .199
/gene="IDS"
/note="G00-120-521"
125. .1777
/gene="IDS"
/EC_number="3.1.6.13"
/db_xref="GDB:G00-120-521"
/product="iduronate 2-sulfatase"
/protein_id="AA063197.1"
/db_xref="GI:184562"
/translation="MPPRTGRGLLGLVSSVCVALGSETQANSTTDALNVLLIIV
DILRPSIGYGVGDKLVSIPNIDQLASHLLFQNAQAVCAVSRLTGRRTDTRL
IDFNSVYRVHAGNFSTIPQVFKNGVYVTVSGVFKHGIGSNHTDDSPYSWSPYPH
SSPKYENTKCRGPDGELHANLLCPVDLVDVPEGLPDKGTEQAQLLEKMTSASP
FFLAVGHKPHIPFRPKFKELPLENITLADPEVDPGLPVPAYNPWMDIRQEDY
QALNISVPYGPDPVQKIRQSFASFVSDTQVGRLLSLDQALQANSTIATFSD
HGWALGEHGEWAKYSNFDVATHVPLIFYPVPGTASUPEAGEKLPFYLDPDFDSASQLME

sig_peptide

CDS

PGROSLVLESLFPTLAGLAGLOVPRCPVPSPHVELCBEGKNLLKHFRPRDLEED
PYLPNCNRELIALYSQYPRPSDIPQWNSDKPSLKDKIKMGYSIRITIDRYITVWVGFNPD
EFLANFSDIIEHAGELYFVDSDBPLQDHNMYNDSQGDLFLQLMP*

gene

mat_peptide

BASE COUNT 553 a 600 c 510 g 634 t

ORIGIN

Query Match 100.0%; Score 2297; DB 6; Length 2297;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 2297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	CGGCTGTGTTGGCAGAGTCTTCATGGTTCGGACAGGAGGTCTCTGTGGCTGCGCGGCG 60
Qy	1	CGGCTGTGTTGGCAGAGTCTTCATGGTTCGGACAGGAGGTCTCTGTGGCTGCGCGGCG 60
Db	61	TGCTAACTGCGCCACCTGCTGCAGCCTGTCCCGCGCTCTGAAGCGCGCGTCAAGC 120
Qy	61	TGCTAACTGCGCCACCTGCTGCAGCCTGTCCCGCGCTCTGAAGCGCGCGTCAAGC 120
Db	121	CGAAATGCGCGCCACCCCGGACGGCGGAGCCCTTCTGTGGCTGGTCTGTTCTGAGCTC 180
Qy	121	CGAAATGCGCGCCACCCCGGACGGCGGAGCCCTTCTGTGGCTGGTCTGTTCTGAGCTC 180
Db	181	CGTCTGCTGCCCTCGGATCCGAACACGAGCCAACTCGACACAGATGCTCTCAAGCT 240
Qy	181	CGTCTGCTGCCCTCGGATCCGAACACGAGCCAACTCGACACAGATGCTCTCAAGCT 240
Db	241	TCCTTCTCATATCGTGGATGACCTCGCGCCCTCCCTGGGCTGTTATGGGATAAAGCTGT 300
Qy	241	TCCTTCTCATATCGTGGATGACCTCGCGCCCTCCCTGGGCTGTTATGGGATAAAGCTGT 300
Db	301	GAGTCCCAATATATGACCAACTGGGATCCGACATCCGACAGCCCTCTCTCCAGAAATGCTTGC 360
Qy	301	GAGTCCCAATATATGACCAACTGGGATCCGACATCCGACAGCCCTCTCTCCAGAAATGCTTGC 360
Db	361	GCAGCAAGCAGTGTGCGCCCGGAGCGGCTTCTTCTTCTCACTGCGAGGAGACCTGACAC 420
Qy	361	GCAGCAAGCAGTGTGCGCCCGGAGCGGCTTCTTCTTCTCACTGCGAGGAGACCTGACAC 420
Db	421	CACCCGCTGTACGACTTCAACTCTCTACTGGAGGTGCACGCTGGAAACTTCTCCACCAT 480
Qy	421	CACCCGCTGTACGACTTCAACTCTCTACTGGAGGTGCACGCTGGAAACTTCTCCACCAT 480
Db	481	CCCCAGTACTTCAAGGAGATGGCTATGTGACCATGTGCGGTGGGAAAGTCTTTCAACC 540
Qy	481	CCCCAGTACTTCAAGGAGATGGCTATGTGACCATGTGCGGTGGGAAAGTCTTTCAACC 540
Db	541	TGGGATATCTTCAACATACCGATGATTCGCCGTATAGCTGGTCTTTCCACCTTATCA 600
Qy	541	TGGGATATCTTCAACATACCGATGATTCGCCGTATAGCTGGTCTTTCCACCTTATCA 600
Db	601	TCCTTCTCTCAGAGATGAAACACATAGACATGTGAGGGCCAGATGGAGACTCCA 660
Qy	601	TCCTTCTCTCAGAGATGAAACACATAGACATGTGAGGGCCAGATGGAGACTCCA 660
Db	661	TGCCAACCTGCTTTCCTGCTGGATGTGCTGGATGTTCGCCAGGGCAGCTTGCCTGACAA 720
Qy	661	TGCCAACCTGCTTTCCTGCTGGATGTGCTGGATGTTCGCCAGGGCAGCTTGCCTGACAA 720
Db	721	ACAGAGCACTGAGCAAGCCATACAGTGTGGAAAGATGAAACGTCAGCCAGTCCCTTT 780
Qy	721	ACAGAGCACTGAGCAAGCCATACAGTGTGGAAAGATGAAACGTCAGCCAGTCCCTTT 780
Db	781	CTTCTGCTGCTTGGGTATACGCCACATACCTTTCAGATACCCCAAGGAATTTCA 840
Qy	781	CTTCTGCTGCTTGGGTATACGCCACATACCTTTCAGATACCCCAAGGAATTTCA 840
Db	841	GAACTTGTATCCCTTGGAGAACATCACCTTGCGCCCGCATCCCGAGGTCCCTTGATGGCT 900
Qy	841	GAACTTGTATCCCTTGGAGAACATCACCTTGCGCCCGCATCCCGAGGTCCCTTGATGGCT 900

Qy	841	GAAGTTGTATCCCTTGGAGAACATCACCTTGGCCCCCGATCCGAGGTCCCTGATGGCCT 900
Db	901	ACCCCTGTGGCTTACAAACCCCTGGATGGACATCAGGCAACGGGAAAGAGTCCAAAGCCCTT 960
Qy	901	ACCCCTGTGGCTTACAAACCCCTGGATGGACATCAGGCAACGGGAAAGAGTCCAAAGCCCTT 960
Db	961	AAACATCAGTGTCCGTATGGTCCAAATTCCTGTGGACTTTTCAGCGGAAATCCGCCAGAG 1020
Qy	961	AAACATCAGTGTCCGTATGGTCCAAATTCCTGTGGACTTTTCAGCGGAAATCCGCCAGAG 1020
Db	1021	CTACTTTGGCTCTGTGTCTATATTTGGATACACAGGTCGCGCGCTCTTGGTGTCTTTGGA 1080
Qy	1021	CTACTTTGGCTCTGTGTCTATATTTGGATACACAGGTCGCGCGCTCTTGGTGTCTTTGGA 1080
Db	1081	CGATCTTCAAGCTGGCCACAGCACCATTGATTTTACCTCGGATCATGGGTGGGCTCT 1140
Qy	1081	CGATCTTCAAGCTGGCCACAGCACCATTGATTTTACCTCGGATCATGGGTGGGCTCT 1140
Db	1141	AGGTGAACATGGAGATGGCCAAATACAGCAATTTGTATGTGTACCATGTTCCTCCCT 1200
Qy	1141	AGGTGAACATGGAGATGGCCAAATACAGCAATTTGTATGTGTACCATGTTCCTCCCT 1200
Db	1201	GATATTCATGTTCCTGGAAGGAGCGCTTCACTTCGAGGAGGCGGAGAGCTTTTCCC 1260
Qy	1201	GATATTCATGTTCCTGGAAGGAGCGCTTCACTTCGAGGAGGCGGAGAGCTTTTCCC 1260
Db	1261	TTACTCTGACCCCTTTTGTATTCGCGCTCACAGTTGTATGGAGCGAGGCAATCCATGGA 1320
Qy	1261	TTACTCTGACCCCTTTTGTATTCGCGCTCACAGTTGTATGGAGCGAGGCAATCCATGGA 1320
Db	1321	CTTGTGGAACCTTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTCGAGTTTC 1380
Qy	1321	CTTGTGGAACCTTGTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTCGAGTTTC 1380
Db	1381	ACTCGCTCCCGCTTCTCTTCAATTCACGTTGAGTGTGAGAGGCAAGCAATCTTCT 1440
Qy	1381	ACTCGCTCCCGCTTCTCTTCAATTCACGTTGAGTGTGAGAGGCAAGCAATCTTCT 1440
Db	1441	GAAGCAATTTCCATTCGCTGAGTGTGGAAGGAGTCCCTACCTCCCTGGTAAATCCCGTGA 1500
Qy	1441	GAAGCAATTTCCATTCGCTGAGTGTGGAAGGAGTCCCTACCTCCCTGGTAAATCCCGTGA 1500
Db	1501	ACTGATTTGCTATAGCAGTATCCCGGCTTTCAGACATCCCTCAGTGGAAATTTCTCACAA 1560
Qy	1501	ACTGATTTGCTATAGCAGTATCCCGGCTTTCAGACATCCCTCAGTGGAAATTTCTCACAA 1560
Db	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTATTCATACGACCATAGACTATAGGTA 1620
Qy	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTATTCATACGACCATAGACTATAGGTA 1620
Db	1621	TACTGTGGTGGTTCCTCAATCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680
Qy	1621	TACTGTGGTGGTTCCTCAATCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680
Db	1681	AGGGAACCTGATTTTGTGGATTCGACCATTCAGGATCACAAATATATTAATGATTC 1740
Qy	1681	AGGGAACCTGATTTTGTGGATTCGACCATTCAGGATCACAAATATATTAATGATTC 1740
Db	1741	CCAAGTGGAGATCTTTTCCAGTTGTGCTTGTGCTTGGTAAACCATGGATGGCAA 1800
Qy	1741	CCAAGTGGAGATCTTTTCCAGTTGTGCTTGTGCTTGGTAAACCATGGATGGCAA 1800
Db	1801	ATGTGATGTGCTCCCTTCCAGTGTGAGAGGAGTGTAGAGCTGGTGGTCTTTGTGATT 1860
Qy	1801	ATGTGATGTGCTCCCTTCCAGTGTGAGAGGAGTGTAGAGCTGGTGGTCTTTGTGATT 1860
Db	1861	ACCCATAATATTTGGAAGCAGCTGAGGCTAGTTAATCCAAACATGATCAATTTGG 1920
Qy	1861	ACCCATAATATTTGGAAGCAGCTGAGGCTAGTTAATCCAAACATGATCAATTTGG 1920
Db	1921	CCTGAGAATATCTAAACAGCAACCTTTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1980
Qy	1921	CCTGAGAATATCTAAACAGCAACCTTTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1980

Db 1981 AATTGGACCAAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTTATTACTG 2040
QY 1981 AATTGGACCAAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTTATTACTG 2040
Db 2041 AATAAATACAAAGCAAAACAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100
QY 2041 AATAAATACAAAGCAAAACAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100
Db 2101 TAATAACCAAAACATACATATATACAAAGAAATACCTTTTCATATTTTGTGAAATTTAGTGC 2160
QY 2101 TAATAACCAAAACATACATATATACAAAGAAATACCTTTTCATATTTTGTGAAATTTAGTGC 2160
Db 2161 ATTTCAAAAAGTAATCATATATACAACTAGGCACCACTAAGTTCCTGATTTATTTTGGT 2220
QY 2161 ATTTCAAAAAGTAATCATATATACAACTAGGCACCACTAAGTTCCTGATTTATTTTGGT 2220
Db 2221 TATAATTTAATATATATCTTATGAGCCCTATATATTTCAAAATATTTATGTTAATCATGTA 2280
QY 2221 TATAATTTAATATATCTTATGAGCCCTATATATTTCAAAATATTTATGTTAATCATGTA 2280
Db 2281 TCCATGTTCTTTTCC 2297
QY 2281 TCCATGTTCTTTTCC 2297

RESULT 2
LOCUS 192327 2297 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 1 from patent US 5728381.
ACCESSION 192327
VERSION 192327.1 GI:3936797
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2297)
AUTHORS Wilson,P.J., Morris,C.Phillip, Anson,D.Stewart, Occhiodoro,T.,
Bielicki,J., Clements,P.Roy and Hopwood,J.Joseph.
TITLE Glycosylation variants of Iduronate 2-sulfatase
JOURNAL Patent: US 5728381-A 1 17-MAR-1998;
FEATURES
source 1. 2297
BASE COUNT 553 a 600 c 510 g 634 t
ORIGIN

Query Match 100.0%; Score 2297; DB 1; Length 2297;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 2297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGGCTGTGTGCGCAGTCTTCATGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC 60
QY 1 CGGCTGTGTGCGCAGTCTTCATGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC 60
Db 61 TCGTAACGCGCCACTGCTGAGCCTGTCCCGCGCGTCTGAAGCGCGCGCTCGAAGC 120
QY 61 TCGTAACGCGCCACTGCTGAGCCTGTCCCGCGCGTCTGAAGCGCGCGCTCGAAGC 120
Db 121 CGAAATCGCGCCACCGCGGCGGAGGCTTCTCTGGCTGGGTCTGGTTCTGAGCTC 180
QY 121 CGAAATCGCGCCACCGCGGCGGAGGCTTCTCTGGCTGGGTCTGGTTCTGAGCTC 180
Db 181 CGTCTCGCTCGCCCTCGGATCGGAACGAGGCGCAACTCGACCACAGATGCTCTGAACGT 240
QY 181 CGTCTCGCTCGCCCTCGGATCGGAACGAGGCGCAACTCGACCACAGATGCTCTGAACGT 240
Db 241 TCTTCTCATCATCGTGGATGACTGCGGCCCTCCCTGGGCTGTTATGGGATAAAGCTGGT 300
QY 241 TCTTCTCATCATCGTGGATGACTGCGGCCCTCCCTGGGCTGTTATGGGATAAAGCTGGT 300
Db 301 GAGGTCCCAATATTTGACCAACTGCGATCCGACGCTTCTTCCAGAAATGCCCTTTC 360
QY 301 GAGGTCCCAATATTTGACCAACTGCGATCCGACGCTTCTTCCAGAAATGCCCTTTC 360

Db 361 GCAGCAAGCAGTGTGCGCCCGAGCGGTTTCTTCTCCTCACTGCAGGAGACCTGACAC 420
QY 361 GCAGCAAGCAGTGTGCGCCCGAGCGGTTTCTTCTCCTCACTGCAGGAGACCTGACAC 420
Db 421 CACCGCGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACATTTCTCCACCAT 480
QY 421 CACCGCGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACATTTCTCCACCAT 480
Db 481 CCCCCAGTACTTCAAGGAGAAATGCTATGACACCATGTCGTTGGGAAAAAGTCTTTTCACCC 540
QY 481 CCCCCAGTACTTCAAGGAGAAATGCTATGACACCATGTCGTTGGGAAAAAGTCTTTTCACCC 540
Db 541 TGGGATATCTTCTAACCACATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCA 600
QY 541 TGGGATATCTTCTAACCACATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCA 600
Db 601 TCCTTCTCTGAGAAAGTATGAAAACACTAAGACATGTGCGAGGGCCAGATGGAGAACTCCA 660
QY 601 TCCTTCTCTGAGAAAGTATGAAAACACTAAGACATGTGCGAGGGCCAGATGGAGAACTCCA 660
Db 661 TGGCAACCTGCTTTGCGCTGTGGATGTGCTGGATGTTCGCCGAGGACCTTTGCCGTGACAA 720
QY 661 TGGCAACCTGCTTTGCGCTGTGGATGTGCTGGATGTTCGCCGAGGACCTTTGCCGTGACAA 720
Db 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAGATGAAAACGTCAGCCAGTCTCTT 780
QY 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAGATGAAAACGTCAGCCAGTCTCTT 780
Db 781 CTTCCTGCGCTGTGGGTATCATAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA 840
QY 781 CTTCCTGCGCTGTGGGTATCATAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA 840
Db 841 GAAGTTGTATCCCTTGGAGAACATCACCTGCGCCCGCATCCCGAGGTCCCTGATGGCCT 900
QY 841 GAAGTTGTATCCCTTGGAGAACATCACCTGCGCCCGCATCCCGAGGTCCCTGATGGCCT 900
Db 901 ACCCCCTGTGGCTTACAAACCTCTGGATGGAGATCAGGCAACGGGAAGAGCTTCAAGCCTT 960
QY 901 ACCCCCTGTGGCTTACAAACCTCTGGATGGAGATCAGGCAACGGGAAGAGCTTCAAGCCTT 960
Db 961 AATCATCAGTGTGCGCTATGTTCCAAATTCCTGTGAGCTTTTCAGCGGAAAAATCCGCCAG 1020
QY 961 AATCATCAGTGTGCGCTATGTTCCAAATTCCTGTGAGCTTTTCAGCGGAAAAATCCGCCAG 1020
Db 1021 CTACTTTGCGCTCTGTGTCATATTTGGATACACAGTGTGGCGCGCTCTTGGAGTCTTTGGA 1080
QY 1021 CTACTTTGCGCTCTGTGTCATATTTGGATACACAGTGTGGCGCGCTCTTGGAGTCTTTGGA 1080
Db 1081 CGATCTTTCAGTGTGCGCCCAACAGCACCATCATTTGCAATTTACCTCGGATCATGGTGGCTCT 1140
QY 1081 CGATCTTTCAGTGTGCGCCCAACAGCACCATCATTTGCAATTTACCTCGGATCATGGTGGCTCT 1140
Db 1141 AGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGCTACCCATGTTCCCT 1200
QY 1141 AGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGCTACCCATGTTCCCT 1200
Db 1201 GATATTTCTATGTTCTGSAAGGCGCTTCACTTCCGAGGCGAGGAGAGCTTTTCCC 1260
QY 1201 GATATTTCTATGTTCTGSAAGGCGCTTCACTTCCGAGGCGAGGAGAGCTTTTCCC 1260
Db 1261 TTACTCGACCTTTTGTATTCGCGCTCACAGTTGATGGAGCCAGGAGCAATCCATGGA 1320
QY 1261 TTACTCGACCTTTTGTATTCGCGCTCACAGTTGATGGAGCCAGGAGCAATCCATGGA 1320
Db 1321 CTTTGTGGAACCTGTGCTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCC 1380
QY 1321 CTTTGTGGAACCTGTGCTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCC 1380
Db 1381 ACCTCGCTGCCCTTCTCTCATTTTCAGTTGAGCTGTGCAGAGAGGCAAGACCTTCT 1440
QY 1381 ACCTCGCTGCCCTTCTCTCATTTTCAGTTGAGCTGTGCAGAGAGGCAAGACCTTCT 1440

Db	1441	GAAGCATTTTCGATTCCTGCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA	1500
Qy	1441	GAAGCATTTTCGATTCCTGCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA	1500
Db	1501	ACTGATTCCTTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGACAA	1560
Qy	1501	ACTGATTCCTTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGACAA	1560
Db	1561	CCGAGTTTAAAGATATAAAGATCATGGGCTATTCCATACGCCACATAGACTATAGSTA	1620
Qy	1561	CCGAGTTTAAAGATATAAAGATCATGGGCTATTCCATACGCCACATAGACTATAGSTA	1620
Db	1621	TACTGTGTGGGTGGCTTCATCTCAGTGAATTTCTAGCTAACTTTCTGACATCCATGC	1680
Qy	1621	TACTGTGTGGGTGGCTTCATCTCAGTGAATTTCTAGCTAACTTTCTGACATCCATGC	1680
Db	1681	AGGGAACTGTATTTTGGGATTCGACCCATTCGAGGATCAATATGTATAAATGATTC	1740
Qy	1681	AGGGAACTGTATTTTGGGATTCGACCCATTCGAGGATCAATATGTATAAATGATTC	1740
Db	1741	CCAGGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTCTTTTGCACCAATGGATGGCAA	1800
Qy	1741	CCAGGTGGAGATCTTTTCCAGTTGTTGATGCTTTTGCACCAATGGATGGCAA	1800
Db	1801	ATGTGATGTGCTCCCTTCCAGCTGTGTGAGAGGAGTGTAGCTGTGCTTTGTGATT	1860
Qy	1801	ATGTGATGTGCTCCCTTCCAGCTGTGTGAGAGGAGTGTAGCTGTGCTTTGTGATT	1860
Db	1861	ACCCATATATTGGAAGCAGCTGAGGCTAGTTAATCCAAATGCATCAACAATTTGG	1920
Qy	1861	ACCCATATATTGGAAGCAGCTGAGGCTAGTTAATCCAAATGCATCAACAATTTGG	1920
Db	1921	CCTGAGATATGTAAACAGCCAAACCTTTTCTGTTAGTCTTTTATAAATTTATAATTTGT	1980
Qy	1921	CCTGAGATATGTAAACAGCCAAACCTTTTCTGTTAGTCTTTTATAAATTTATAATTTGT	1980
Db	1981	AATTGGACCACTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	2040
Qy	1981	AATTGGACCACTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	2040
Db	2041	AATAAATACAAGCAAACTCAAGTTATGTCTATCTTTTGGATAGCAAGACCATACA	2100
Qy	2041	AATAAATACAAGCAAACTCAAGTTATGTCTATCTTTTGGATAGCAAGACCATACA	2100
Db	2101	TAATAACCAACATACATATATACAAAGAACTCTTCAATATTTTGGAAATTTAGTGC	2160
Qy	2101	TAATAACCAACATACATATATACAAAGAACTCTTCAATATTTTGGAAATTTAGTGC	2160
Db	2161	ATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTAAGTTCCCTGATTATTTTGT	2220
Qy	2161	ATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTAAGTTCCCTGATTATTTTGT	2220
Db	2221	TATAATTTAATAATATATCTTATGAGCCCTATATATTCAAATATATATTTAAACATGAA	2280
Qy	2221	TATAATTTAATAATATATCTTATGAGCCCTATATATTCAAATATATATTTAAACATGAA	2280
Db	2281	TCCATGTTCTTTTCC 2297	
Qy	2281	TCCATGTTCTTTTCC 2297	
RESULT	3	AR025358 2297 bp DNA PAT 05-DEC-1998	
LOCUS		Sequence 1 from patent US 5798239.	
DEFINITION		AR025358	
ACCESSION		AR025358.1 GI:3977986	
VERSION		Unknown.	
KEYWORDS		Unknown.	
SOURCE		Unclassified.	
ORGANISM		1 (bases 1 to 2297)	
REFERENCE		Wilson,P.J., Morris,C.,Phillip, Anson,D.,Stewart, Occhiodoro,T.,	
AUTHORS		Bielicki,J., Clements,P. Roy and Hopwood,J. Joseph.	

TITLE	Glycosylation variants of iduronate 2-sulfatase		
JOURNAL	Patent: US 5798239-A 1 25-AUG-1998;		
FEATURES	Location/Qualifiers		
	1..2297		
source	/organism="unknown"		
BASE COUNT	553 a	600 c	510 g 634 t
ORIGIN			
Query Match	100.0%;	Score 2297;	DB 1; Length 2297;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;	
Matches 2297;	Conservative	0; Mismatches	0; Indels 0; Gaps 0;
Db	1	CGGCTGTGTGGCGAGCTTTCATGTGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC	60
Qy	1	CGGCTGTGTGGCGAGCTTTCATGTGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC	60
Db	61	TGCTAACTGGCGACCTGTCTGAGCCTGTCCCGCGCGCTCTGAAGCGCGCGCTCGAAGC	120
Qy	61	TGCTAACTGGCGACCTGTCTGAGCCTGTCCCGCGCGCTCTGAAGCGCGCGCTCGAAGC	120
Db	121	CGAAATGCCGCCACCCGCGGACCGCGGAGCGCTTCTCTGGCTGGTCTGTCTGAGCTC	180
Qy	121	CGAAATGCCGCCACCCGCGGAGCGCGGAGCGCTTCTCTGGCTGGTCTGTCTGAGCTC	180
Db	181	CGTCTGCGTGGCGCTCGGATCCGAAACGCGAGGCCAACTCGACACAGATGCTCTGAACGT	240
Qy	181	CGTCTGCGTGGCGCTCGGATCCGAAACGCGAGGCCAACTCGACACAGATGCTCTGAACGT	240
Db	241	TCTTCTCATCATCTGTGGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGATAGCTGGT	300
Qy	241	TCTTCTCATCATCTGTGGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGATAGCTGGT	300
Db	301	GAGTCCCCCAATATTGACCAACTGGCATCCACAGCCTCCTCTCCAGAAATSCCTTTGC	360
Qy	301	GAGTCCCCCAATATTGACCAACTGGCATCCACAGCCTCCTCTCCAGAAATSCCTTTGC	360
Db	361	GCAGCAAGCAGTGTGGCGCCCGGAGCGCTTTCTTCTCACTGCGAGGAGACTGACAC	420
Qy	361	GCAGCAAGCAGTGTGGCGCCCGGAGCGCTTTCTTCTCACTGCGAGGAGACTGACAC	420
Db	421	CACCGCGCTGTACGACTTCAACTCTTCTGAGGGTGCACGCTGGAACCTTCTCCACCAT	480
Qy	421	CACCGCGCTGTACGACTTCAACTCTTCTGAGGGTGCACGCTGGAACCTTCTCCACCAT	480
Db	481	CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGGAAAAGTCTTTTCCACC	540
Qy	481	CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGGAAAAGTCTTTTCCACC	540
Db	541	TGGGATATCTTCTAACATACCGATGATTCTCCGATAGCTGTCTTTCCACCTTATCA	600
Qy	541	TGGGATATCTTCTAACATACCGATGATTCTCCGATAGCTGTCTTTCCACCTTATCA	600
Db	601	TCCTCTCTCGAAGATATGAAAACACTAAGACATGTGAGGGGCCAGATGGAGAACTCCA	660
Qy	601	TCCTCTCTCGAAGATATGAAAACACTAAGACATGTGAGGGGCCAGATGGAGAACTCCA	660
Db	661	TGCGAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCCGAGGGCACCTTGCCTGACAA	720
Qy	661	TGCGAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCCGAGGGCACCTTGCCTGACAA	720
Db	721	ACAGAGCACTGAGCAAGCCATACAGTTGTTGAAAAGATGAAAACCTGACGAGTCTTTT	780
Qy	721	ACAGAGCACTGAGCAAGCCATACAGTTGTTGAAAAGATGAAAACCTGACGAGTCTTTT	780
Db	781	CTTCTGCGCGTTGGGTATATAAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA	840
Qy	781	CTTCTGCGCGTTGGGTATATAAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA	840
Db	841	GAAGTTGTATCCCTTGGAGAACATCACCTGGCCCCCGATCCCGAGGTCCCTGATGGCCT	900
Qy	841	GAAGTTGTATCCCTTGGAGAACATCACCTGGCCCCCGATCCCGAGGTCCCTGATGGCCT	900
Db	901	ACCCCTGTGGCTTACACCCCTGGATGGACATCAGGCAACGGGAAGACGCTCCAGCCTT	960

Qy	1981	AATTGGACCAAGT	TTTTTTTTTAAATTTCCCTCTTTTTTAAACAACAGTTACGCGCTTATTCTAGT	204
Db	2041	AATAAATACAAGCAACAACAACTCAAGT	TATGTCTATACATACCTTTGGATAGCAAGACCATACA	2100
Qy	2041	AATAAATACAAGCAACAACAACTCAAGT	TATGTCTATACATACCTTTGGATAGCAAGACCATACA	2100
Db	2101	TAATAACCAACATACATATATACACAAGAATAC	TTTCATTATTTGTGGAAATTTAGTGCC	2160
Qy	2101	TAATAACCAACATACATATATACACAAGAATAC	TTTCATTATTTGTGGAAATTTAGTGCC	2160
Db	2161	ATTTTCAAAAAGTAAATCATATATCAAACTAGG	CACCACACTTAAGTTGCTTCGATATTTTTCGT	2220
Qy	2161	ATTTTCAAAAAGTAAATCATATATCAAACTAGG	CACCACACTTAAGTTGCTTCGATATTTTTCGT	2220
Db	2221	TATAATTTAATAATATATCTTATGAGCCCTATATAT	TTCAAAAATATATGTTAAACATGTAA	2280
Qy	2221	TATAATTTAATAATATATCTTATGAGCCCTATATAT	TTCAAAAATATATGTTAAACATGTAA	2280
Db	2281	TCCCATGTTCTTTTTTCC	2297	
Qy	2281	TCCCATGTTCTTTTTTCC	2297	
RESULT	4			
LOCUS		HUMIDSALT	1314 bp mRNA	
DEFINITION		Homo sapiens iduronate-2-sulphatase (IDS) mRNA, complete cds.		11-MAR-1996
ACCESSION		L40586		
VERSION		L40586.1	GI:1222629	
KEYWORDS		Iduronate-2-sulphatase.		
SOURCE		Homo sapiens		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS		1 (bases 1 to 1314)		
TITLE		Malmgren,H., Carlberg,B.M., Pettersson,U. and Bondeson,M.L. Identification of an alternative transcript from the human iduronate-2-sulfatase (IDS) gene		
JOURNAL		Genomics 29 (1), 291-293 (1995)		
MEDLINE		96079126		
FEATURES		Location/Qualifiers		
source		1. .1314		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="23.2"		
		/tissue_type="lymphocyte"		
gene		1. .1291		
		/gene="IDS"		
exon		1. .230		
		/gene="IDS"		
		/note="G00-120-521"		
		/number=1		
CDS		128. .1159		
		/gene="IDS"		
		/codon_start=1		
		/db_xref="GBB:G00-120-521"		
		/product="iduronate-2-sulphatase"		
		/protein_id="AA92014.1"		
		/db_xref="GI:1222630"		
		/translation="MPPPTGRGLWLGLVSSVCVALGSETQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEKMKTFPLAVGYHKPIFPYRPFKEFKLYPLENTILAPDPEVDPGLPPVAYNPWMDIRQFLQANISVPYPIPVDFQKIROISYFASVSYLDTVQVGRLLSALDQLQANSTTDAINVLDLRLSLGCGYDKLVRSPNIDQLASHLSLLFQNAFAQCAVCASRVSLTGRREDFNSVWRVHAGNST IPOYFKENGYVTMSVKYFHPGISNHTDDSPYSWSPSSPKYENTKTCRGPDELHANLLCPVDLDVPEGLPKDQSTEOAIQLLEK		


```
/tissue_lib="lambda Charon 35, ATCC No. LAOXN101"
/map="Xq27.3-q28"
join(L13321.1:1..536,L13322.1:1..426,L13323.1:1..388,
L13324.1:1..366,L13325.1:1..490,L13326.1:1..379,
L13327.1:1..319,L13328.1:1..404,1..1120)
/gene="IDS"
CDS
join(L13321.1:332..434,L13322.1:157..293,L13323.1:82..259,
L13324.1:131..219,L13325.1:125..325,L13326.1:188..258,
L13327.1:99..225,L13328.1:129..302,128..600)
/gene="IDS"
/codon_start=1
/product="iduronate-2-sulfatase"
/protein_id="AAJ16877.1"
/db_xref="GI:405205"
/translacion="MPPRTGRLLWLGVLSSVCVALGSETQANSTFDALNLLIIV
DDLRPSGCDKLVIRPNIDQLASHLLFQNAFAQAVCAPSRVSFLTGRPDTRL
YDFNSWRVHAGNESTIPOYFKENGIVTMSGVKFPHPGISNHTDDSPYSFPPHP
SSEKYENTKCRGPDGELHANLCPVDVLDVPECTLPDKOSTEOAIOLLEKWTASAP
FFLAVGYKHPHIPPYKFKQLYPLENITLAPDPEVDPGLPPVAYNPWMDIREDV
QALNISYPGIPVDFOFKIKQSYFASVSYLDTOVGRLLSALDLDLANSTIIAFTSD
HGVALGHEWAKYSNFDVATHVLFIYVPGRTASLPEAGEKLFYLDPDFDSASQLE
PQLGMDLVELSLFPTLAGLAGLVPPRCVPVPSFHVLELREGKLNLLKHFRFROLEE
PYLPNPRELIAVSOYPRPSDIPQWNSDKPSLKDKIKMGYSIRTDIYRYTVWGFNP
EFLANFSDIHAGELYFVDSPLQDHNYNDSQGDLPQLMP"
intron
order(L13328.1:303..404,1..127)
/gene="IDS"
/number=8
/number=9
/number=8
128..>1120
/gene="IDS"
/number=9
/evidence-experimental
BASE COUNT 318 a 232 c 212 g 358 t
ORIGIN

Query Match 43.4%; Score 996; DB 6; Length 1120;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 CAGCGAGCAATCCATGGACCTTGTGGAACTGTGTCCTTTTCCACGCTGGAC 184
QY 1302 CAGCGAGCAATCCATGGACCTTGTGGAACTGTGTCCTTTTCCACGCTGGAC 1361
Db 185 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCCTTCATTTCCAGCTGTGCA 244
QY 1362 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCCTTCATTTCCAGCTGTGCA 1421
Db 245 GAGAGGCAAGAACTTCTGAGCAATTTTCGATCCGCTGACTTGAAGAGGATCGTACC 304
QY 1422 GAGAGGCAAGAACTTCTGAGCAATTTTCGATCCGCTGACTTGAAGAGGATCGTACC 1481
Db 305 TCCCTGGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 364
QY 1482 TCCCTGGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
Db 365 CTGAGTGGAACTTCGACAGCGAGTTCGCTATAGAGATATAAGATCATGGGCTATTCATAC 424
QY 1542 CTGAGTGGAACTTCGACAGCGAGTTCGCTATAGAGATATAAGATCATGGGCTATTCATAC 1601
Db 425 GCACCATAGACTATAGGTATAGTGTGGGTGGTTCGCTCAATCCGATGAATTTCTAGCTA 484
QY 1602 GCACCATAGACTATAGGTATAGTGTGGGTGGTTCGCTCAATCCGATGAATTTCTAGCTA 1661
Db 485 ACTTTTCTGACATCCATCGAGGAACTGATTTTGTGGATCTCGACCATTCGAGGATC 544
QY 1662 ACTTTTCTGACATCCATCGAGGAACTGATTTTGTGGATCTCGACCATTCGAGGATC 1721
Db 545 ACAATATGTAATGATTCACCAAGTGGAGATCTTTTCCAGTGTGATGCTTCAGTTT 604
QY 1722 ACAATATGTAATGATTCACCAAGTGGAGATCTTTTCCAGTGTGATGCTTCAGTTT 1781
Db 605 TGCCAACCATGGATGGCAAAATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 664
QY 1782 TGCCAACCATGGATGGCAAAATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 1841
```

```
Db 665 AGCTGTCGCTTTTGTGATTACCCATTAATATTGGAAGCAGCCTGAGGCTAGTTAATCCAA 724
QY 1842 AGCTGTCGCTTTTGTGATTACCCATTAATATTGGAAGCAGCCTGAGGCTAGTTAATCCAA 1901
Db 725 ACATGCATCAACAATTTGGCCCTGAGATATGTAACAGCCAAACCTTTTCGTTAGTCTTT 784
QY 1902 ACATGCATCAACAATTTGGCCCTGAGATATGTAACAGCCAAACCTTTTCGTTAGTCTTT 1961
Db 785 ATTAATAATTTATAATTTGTAATTTGGACCAGTCTTTTAAATTTCCCTCTTTTAAAC 844
QY 1962 ATTAATAATTTATAATTTGTAATTTGGACCAGTCTTTTAAATTTCCCTCTTTTAAAC 2021
Db 845 AGTTAGCGCTTTATTTACTGAATAAATACAAAGCAAAACAACTCAAGTTATGTGCATACCTT 904
QY 2022 AGTTAGCGCTTTATTTACTGAATAAATACAAAGCAAAACAACTCAAGTTATGTGCATACCTT 2081
Db 905 TGGATACGAGAACCATACATATAAACAACAATATATACACAAGAAATACTTTCAT 964
QY 2082 TGGATACGAGAACCATACATATAAACAACAATATATACACAAGAAATACTTTCAT 2141
Db 965 TATTGTGCAATTTACTGCAATTTCAAAAAGTAATATATATCAAACTAGGCACACACTA 1024
QY 2142 TATTGTGCAATTTACTGCAATTTCAAAAAGTAATATATATCAAACTAGGCACACACTA 2201
Db 1025 AGTTCTCTGATTTTGTGTTTAAATTTAAATATATATCTTATGAGCCCTATATATCAAA 1084
QY 2202 AGTTCTCTGATTTTGTGTTTAAATTTAAATATATATCTTATGAGCCCTATATATCAAA 2261
Db 1085 ATATTATGTTAAATGTAATCCATGTTCTTTTTC 1120
QY 2262 ATATTATGTTAAATGTAATCCATGTTCTTTTTC 2297
```

```
RESULT 7
LOCUS AR025360 4428 bp DNA PAT 05-DEC-1998
DEFINITION Sequence 6 from patent US 5798239.
ACCESSION AR025360
VERSION AR025360.1 GI:3977988
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4428)
AUTHORS Wilson,P.J., Morris,C.Phillip, Anson,D.Stewart, Occhiodoro,T.,
Bielicki,J., Clements,P.Roy and Hopwood,J.Joseph.
TITLE Glycosylation variants of iduronate 2-sulfatase
JOURNAL Patent: US 5798239-A 6 25-AUG-1998;
FEATURES Location/Qualifiers
source 1..4428
BASE COUNT 1033 a 1064 c 1042 g 1285 t 4 others
ORIGIN

Query Match 43.4%; Score 996; DB 1; Length 4428;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3433 CAGCGAGCAATCCATGGACCTTGTGGAACCTTGTGTCCTTTTCCACGCTGGCTGAC 3492
QY 1302 CAGCGAGCAATCCATGGACCTTGTGGAACCTTGTGTCCTTTTCCACGCTGGCTGAC 1361
Db 3493 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCCTTCATTTACGTTGAGCTGTGCA 3552
QY 1362 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCCTTCATTTACGTTGAGCTGTGCA 1421
Db 3553 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGGAAAGAGTCCGTACC 3612
QY 1422 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGGAAAGAGTCCGTACC 1481
Db 3613 TCCCTGGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 3672
QY 1482 TCCCTGGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
```

Db 3673 CTCAGTGAATTCGACAGCGAGCTTTAAAGATATATAAGATCATGGGCTATTCATAC 3732
|||||
Qy 1542 CTCAGTGAATTCGACAGCGAGCTTTAAAGATATATAAGATCATGGGCTATTCATAC 1601
Db 3733 GCACCATAGACTATAGGTATAGTGTGGGTTGGCTTCAATCCTGAAGAATTTCTAGCTA 3792
Qy 1602 GCACCATAGACTATAGGTATAGTGTGGGTTGGCTTCAATCCTGAAGAATTTCTAGCTA 1661
Db 3793 ACTTTCTGACATCCATCGAGGGGAACCTGATTTTGTGGATCTGTGACCCATTCGAGGATC 3852
|||||
Qy 1662 ACTTTCTGACATCCATCGAGGGGAACCTGATTTTGTGGATCTGTGACCCATTCGAGGATC 1721
Db 3853 ACAATATGTAATGATTTCCCAAGGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 3912
|||||
Qy 1722 ACAATATGTAATGATTTCCCAAGGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 1781
Db 3913 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 3972
Qy 1782 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 1841
Db 3973 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTTGAGGGCTAGTTAATCCAA 4032
|||||
Qy 1842 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTTGAGGGCTAGTTAATCCAA 1901
Db 4033 ACATGCATCAACAATTTGGCCTGAGAATATGTAACAGGCCAAACCTTTTCGTTTAGTCTTT 4092
|||||
Qy 1902 ACATGCATCAACAATTTGGCCTGAGAATATGTAACAGGCCAAACCTTTTCGTTTAGTCTTT 1961
Db 4093 ATTAATAATTTAATTTGGTAATTTGGACCAAGTTTGTGTTTAAATTTCCCTCTTTTAAAC 4152
|||||
Qy 1962 ATTAATAATTTAATTTGGTAATTTGGACCAAGTTTGTGTTTAAATTTCCCTCTTTTAAAC 2021
Db 4153 AGTTACGGCTTATTTACTGAAATTAATAACAAAGCAAACTCAAGTTATGTCATACCTT 4212
Qy 2022 AGTTACGGCTTATTTACTGAAATTAATAACAAAGCAAACTCAAGTTATGTCATACCTT 2081
Db 4213 TGGATACGAAGACCATACATATAACCAACATACATTTATACACAAAGAAATTTTCAT 4272
Qy 2082 TGGATACGAAGACCATACATATAACCAACATACATTTATACACAAAGAAATTTTCAT 2141
Db 4273 TATTTGTGGAATTTAGTCATTTCAAAAGTAATCATATATCAAACTAGGCACACACTA 4332
|||||
Qy 2142 TATTTGTGGAATTTAGTCATTTCAAAAGTAATCATATATCAAACTAGGCACACACTA 2201
Db 4333 AGTTCCCTGATTTTGTGTTTAAATTAATATATCTTATGAGCCCTATATATCAAA 4392
Qy 2202 AGTTCCCTGATTTTGTGTTTAAATTAATATATCTTATGAGCCCTATATATCAAA 2261

RESULT 8 192329 4428 bp DNA PAT 01-DEC-1998
LOCUS Sequence 6 from patent US 5728381.
DEFINITION
ACCESSION 192329
VERSION 192329.1 GI:3936799
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4428)
AUTHORS Wilson, P.J., Morris, C. Phillip, Anson, D. Stewart, Occhiodoro, T.,
Bielicki, J., Clements, P. Roy and Hopwood, J. Joseph.
TITLE Glycosylation variants of iduronate 2-sulfatase
JOURNAL Patent: US 5728381-A 6 17-MAR-1998;
FEATURES Location/Qualifiers
source 1..4428
/organism="unknown"
BASE COUNT 1033 a 1064 c 1042 g 1285 t 4 others
ORIGIN

Query Match 43.4%; Score 996; DB 1; Length 4428;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3433 CAGCGAGCAATCCATGAGACCTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGAC 3492
Qy 1302 CAGCGAGCAATCCATGAGACCTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGAC 1361
Db 3493 TTGCGAGACTGCGAGTTCCACCTGCGCTGCGCCCGTTCCTTCAATTTCAAGTTGAGCTGTGCA 3552
Qy 1362 TTGCGAGACTGCGAGTTCCACCTGCGCTGCGCCCGTTCCTTCAATTTCAAGTTGAGCTGTGCA 1421
Db 3553 GAGAAGCAAGAACCTTCTGAAGCATTTTCGATTTCCGTTGACTTGGAGAGAGATCCGATCC 3612
Qy 1422 GAGAAGCAAGAACCTTCTGAAGCATTTTCGATTTCCGTTGACTTGGAGAGAGATCCGATCC 1481
Db 3613 TCCTTGTAATCCCGGTGAACCTGATTTGCCCTATAGCCAGTATCCCGGCTTCAGACATCC 3672
Qy 1482 TCCTTGTAATCCCGGTGAACCTGATTTGCCCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
Db 3673 CTCAGTGGAAATTCGACAGCGGAGTTTAAAGATATAAAGATCATGGGCTATTTCCATAC 3732
Qy 1542 CTCAGTGGAAATTCGACAGCGGAGTTTAAAGATATAAAGATCATGGGCTATTTCCATAC 1601
Db 3733 GCACCATAGACTATAGGTATAGTGTGGGTTGGCTTCAATCCCTGATGAATTTCTAGCTA 3792
Qy 1602 GCACCATAGACTATAGGTATAGTGTGGGTTGGCTTCAATCCCTGATGAATTTCTAGCTA 1661
Db 3793 ACTTTCTGACATCCATCGAGGGGAACCTGATTTTGTGGATTTGAGTTTCCAGTTGAGTTT 3852
Qy 1662 ACTTTCTGACATCCATCGAGGGGAACCTGATTTTGTGGATTTGAGTTTCCAGTTGAGTTT 1721
Db 3853 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTTT 3912
Qy 1722 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTTT 1781
Db 3913 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 3972
Qy 1782 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 1841
Db 3973 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 4032
Qy 1842 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 1901
Db 4033 ACATGCATCAACAATTTGGCTGAGAATATGTAACAGCAAACTTTTTCGTTTGTAGTCTTT 4092
Qy 1902 ACATGCATCAACAATTTGGCTGAGAATATGTAACAGCAAACTTTTTCGTTTGTAGTCTTT 1961
Db 4093 ATTAATAATTTAATTTGGTAATTTGGACCAAGTTTGTGTTTAAATTTCCCTCTTTTAAAC 4152
Qy 1962 ATTAATAATTTAATTTGGTAATTTGGACCAAGTTTGTGTTTAAATTTCCCTCTTTTAAAC 2021
Db 4153 AGTTACGGCTTATTTACTGAAATTAATAACAAAGCAAACTCAAGTTATGTCATACCTT 4212
Qy 2022 AGTTACGGCTTATTTACTGAAATTAATAACAAAGCAAACTCAAGTTATGTCATACCTT 2081
Db 4213 TGGATACGAAGACCATACATATAACCAACATACATTTATACACAAAGAAATTTTCAT 4272
Qy 2082 TGGATACGAAGACCATACATATAACCAACATACATTTATACACAAAGAAATTTTCAT 2141
Db 4273 TATTTGTGGAATTTAGTCATTTCAAAAGTAATCATATATCAAACTAGGCACACACTA 4332
Qy 2142 TATTTGTGGAATTTAGTCATTTCAAAAGTAATCATATATCAAACTAGGCACACACTA 2201
Db 4333 AGTTCCCTGATTTTGTGTTTAAATTAATATATCTTATGAGCCCTATATATCAAA 4392
Qy 2202 AGTTCCCTGATTTTGTGTTTAAATTAATATATCTTATGAGCCCTATATATCAAA 2261
Db 4393 ATATTATGTTAAGCATGAATCCCATGTTTCTTTTCC 4428
Qy 2262 ATATTATGTTAAGCATGAATCCCATGTTTCTTTTCC 2297

```

RESULT 9
LOCUS HUMIDS 36845 bp DNA PRI 16-AUG-1994
DEFINITION Homo sapiens iduronate sulphate sulphatase (IDS) gene, complete
          cds.
ACCESSION L35485
VERSION L35485.1 GI:530140
KEYWORDS iduronate sulphate sulphatase.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 36845)
AUTHORS Lu,F., Lu,J., Clingan,R.L., Wentland,M.A., Muzny,D.M., Gu,Y.,
         Nelson,D.L. and Gibbs,R.A.
TITLE Complete DNA sequence of the human iduronate sulphate sulphatase
      (ids) locus
JOURNAL Unpublished (1994)
FEATURES             Location/Qualifiers
     source           1..36845
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="Xg27.3-q28"
     gene             join(1057..1159,1901..2037,2704..2881,5155..5243,
                     7885..8085,9676..9846,15752..15878,19094..19267,
                     22974..23446)
                     /gene="IDS"
     exon             <1057..1159
                     /gene="IDS"
                     /number=1
     CDS              join(1057..1159,1901..2037,2704..2881,5155..5243,
                     7885..8085,9676..9846,15752..15878,19094..19267,
                     22974..23446)
                     /gene="IDS"
                     /codon_start=1
                     /product="iduronate sulphate sulphatase"
                     /protein_id="AAA36037.1"
                     /db_xref="GI:530141"
                     /translation="MPPPTGRGLLGLVSSVVALGSETQANSTTDAINLVLLIIV
                     DLRPSGCGDKIVRSPNIDOLASHSLFQNAFAQAQVAPSRVSFLTGRPDTRL
                     YDENSRYRHNAGNFTIPQYKENGIVTMSGVKVFHGISNHNDDSDSYSWFPPYHP
                     SSEYENKTCRGDGLHNLCPVDLDVPEGLTPDKQTEQAQLLEKMKTSASP
                     FFLAVGHKHIPPRYPKEFKLYPLENITLAPDPEVDPGLPPVAYNPWMDIQRSDV
                     QALNISVPYGPDPVDFQKIQRSYFASVYLDTOVGRLLSALDLDLANSTIIAFTSD
                     HGVALGHEWAKVNSDVAHVPLIEVPCRTASLPEAGEKLPYLPDPDSASOLME
                     PGRSMDLVELVLEPLAGLAGLOVPPRCPVPSEHVELCREGNLLKHFRFRDLEED
                     PYLEGNFRELAYSQYPRSPDIPQWNSDKPSLKDIIKINGYSIRTIDYRTYVWVGPNPD
                     EFLANFSDIHAGELYFVDSPLQDNHMYNDSQGGDLFQLLMP"
                     /gene="IDS"
     intron            1901..2037
                     /number=1
     exon             1901..2037
                     /gene="IDS"
     intron            2038..2703
                     /number=2
                     /gene="IDS"
     exon             2704..2881
                     /number=2
                     /gene="IDS"
     intron            2882..5154
                     /number=3
                     /gene="IDS"
     exon             5155..5243
                     /number=3
                     /gene="IDS"
     intron            5244..7884
                     /number=4
                     /gene="IDS"
     repeat_region     6827..7114
                     /rpt_family="Alu"
     repeat_region     7117..7647
                     /rpt_family="L1nel"

```

```

exon 7885..8085
      /gene="IDS"
      /number=5
intron 8086..9675
      /gene="IDS"
      /number=5
exon 9676..9846
      /gene="IDS"
      /number=6
intron 9847..15751
      /gene="IDS"
      /number=6
exon 15752..15878
      /gene="IDS"
      /number=7
intron 15879..19093
      /gene="IDS"
      /number=7
exon 19094..19267
      /gene="IDS"
      /number=8
intron 19268..22972
      /gene="IDS"
      /number=8
repeat_region 21573..21864
      /rpt_family="Alu"
repeat_region 21867..22051
      /rpt_family="L1nel"
repeat_region 22096..22267
      /rpt_family="L1nel"
exon 22974..>23446
      /gene="IDS"
      /number=9
repeat_region 25413..25686
      /rpt_family="Alu"
repeat_region 26878..27162
      /rpt_family="Alu"
repeat_region 29858..30098
      /rpt_family="Alu"
repeat_region 30767..31068
      /rpt_family="Alu"
repeat_region 33970..34090
      /rpt_family="Alu"
BASE COUNT 9117 a 8696 c 8303 g 10729 t
ORIGIN

Query Match 43.3%; Score 995; DB 6; Length 36845;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22971 CAGGCAGGCAATCCATGGACCTTGTGGAACTTGTGCTCTCTTTTCCACACCTGGCTGGAC 23030
QY 1302 CAGGCAGGCAATCCATGGACCTTGTGGAACTTGTGCTCTCTTTTCCACACCTGGCTGGAC 1361
Db 23031 TTGCAGSACTGCAGGTTCCACCTCGCTGCCCGCTTCCATTCACGTTGAGCTGTGCA 23090
QY 1362 TTGCAGSACTGCAGGTTCCACCTCGCTGCCCGCTTCCATTCACGTTGAGCTGTGCA 1421
Db 23091 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGAAGAGAGATCCCGTACC 23150
QY 1422 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGAAGAGAGATCCCGTACC 1481
Db 23151 TCCCTGGTAATCCCGCTGAACCTGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCC 23210
QY 1482 TCCCTGGTAATCCCGCTGAACCTGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCC 1541
Db 23211 CTCAGTGGAAATCTGACAGCCGAGTTTAAAGATATAAGATCATGGGCTATTTCATAC 23270
QY 1542 CTCAGTGGAAATCTGACAGCCGAGTTTAAAGATATAAGATCATGGGCTATTTCATAC 1601
Db 23271 GCACCATAGACTATAGTATAGTCTGTGGTTTGGCTTCAATCCTGATGAATTTCTAGCTA 23330
QY 1602 GCACCATAGACTATAGTATAGTCTGTGGTTTGGCTTCAATCCTGATGAATTTCTAGCTA 1661

```

Db 23331 ACTTTCTGACATCCATGCAGGGAACCTGATTTTGTGATTTCTGACCCATTGCAGATC 23390
|||||
Qy 1662 ACTTTCTGACATCCATGCAGGGAACCTGATTTTGTGATTTCTGACCCATTGCAGATC 1721
|||||
Db 23391 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTTGTTGATGCCCTTGAGTTT 23450
|||||
Qy 1722 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTTGTTGATGCCCTTGAGTTT 1781
|||||
Db 23451 TGCCAACCATGGATGGCAAAATGATGTCGTCCTCCAGCTGGTGAGAGGAGGATAG 23510
|||||
Qy 1782 TGCCAACCATGGATGGCAAAATGATGTCGTCCTCCAGCTGGTGAGAGGAGGATAG 1841
|||||
Db 23511 AGCTGGTGGTTTGTGATTACCCATAATATTGGAAGCAGCCCTGAGGGCTAGTTAATCCAA 23570
|||||
Qy 1842 AGCTGGTGGTTTGTGATTACCCATAATATTGGAAGCAGCCCTGAGGGCTAGTTAATCCAA 1901
|||||
Db 23571 ACATGCATCAACAATTTGGCCTGGAATATGTAACAGCAACCTTTTCGTTTAGTCTTT 23630
|||||
Qy 1902 ACATGCATCAACAATTTGGCCTGGAATATGTAACAGCAACCTTTTCGTTTAGTCTTT 1961
|||||
Db 23631 ATTAATAATTTAATTTGGTAATTTGACACAGTTTATTTTAAATTCCTCTTTTAAAC 23690
|||||
Qy 1962 ATTAATAATTTAATTTGGTAATTTGACACAGTTTATTTTAAATTCCTCTTTTAAAC 2021
|||||
Db 23691 AGTTACGCTTATTTACTGTAATAATCAAAAGCAACCAACTCAAGTTATGTCATACCTT 23750
|||||
Qy 2022 AGTTACGCTTATTTACTGTAATAATCAAAAGCAACCAACTCAAGTTATGTCATACCTT 2081
|||||
Db 23751 TGGATAGCAAGACCATACATAATAACCAACATACATATACAAAGAAATCTTTCAT 23810
|||||
Qy 2082 TGGATAGCAAGACCATACATAATAACCAACATACATATACAAAGAAATCTTTCAT 2141
|||||
Db 23811 TATTGTGGAATTTAGTGCATTTCAAAAGTAATCATATATCAAACTAGGCACCACTA 23870
|||||
Qy 2142 TATTGTGGAATTTAGTGCATTTCAAAAGTAATCATATATCAAACTAGGCACCACTA 2201
|||||
Db 23871 AGTTCCCTGATTTTGTGTTTATATTTAATATATATCTTATGAGCCCTATATATCAA 23930
|||||
Qy 2202 AGTTCCCTGATTTTGTGTTTATATTTAATATATATCTTATGAGCCCTATATATCAA 2261
|||||
Db 23931 ATATTGTTTAAACATGTAATCATCTTCTTTTTC 23965
|||||
Qy 2262 ATATTGTTTAAACATGTAATCATCTTCTTTTTC 2296
|||||
RESULT 10
LOCUS AF011889 326663 bp DNA PRI 17-NOV-1998
DEFINITION Human Xq28 cosmids U126G1, U142F2, U69B6, U145C10, U169A5, U84H1,
U24D12, U80A7, U153E6, L35485, and R7-163A8 containing iduronate
2-sulfatase gene and pseudogene, complete sequence.
ACCESSION AF011889 U78816 U65082 L39643 L43581 L35485
VERSION AF011889.1 GI:2335186
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 195603 to 325470)
AUTHORS Timms,K.M., Lu,F., Shen,Y., Pleron,C.A., Muzny,D.M., Gu,Y.,
Nelson,D.L. and Gibbs,R.A.
TITLE 130 kb of DNA sequence reveals two new genes and a regional
duplication distal to the human iduronate-2-sulfate sulfatase locus
JOURNAL Genome Res. 5 (1), 71-78 (1995)
MEDLINE 96352905
REFERENCE 2 (bases 120609 to 326663)
AUTHORS Timms,K.M., Bondeson,M.L., Ansari-Lari,M.A., Lagerstedt,K.,
Muzny,D.M., Dugan-Rocha,S.P., Nelson,D.L., Pettersson,U. and
Gibbs,R.A.
TITLE Molecular and phenotypic variation in patients with severe Hunter
syndrome
JOURNAL Hum. Mol. Genet. 6 (3), 479-486 (1997)
MEDLINE 97227295

REFERENCE
AUTHORS

3 (bases 27820 to 69003)
Kerker,S.C., Matsunami,N., Plaetke,R., Albertsen,H., Ballard,L.,
Melis,R., Lawrence,E., Moore,M., Holik,P.R., Carlson,M., Zhao,X.,
Robertson,M., Bradley,P., Elsner,T., Tingey,A., Lalouel,J.-M. and
White,R.

Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome

JOURNAL
REFERENCE
AUTHORS

4 (bases 1 to 326663)
Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Gonzalez,O.,
Haywood,M., Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L.,
Shen,H., Worley,K., Chen,E., Forcum,J., Aronson,A.D., Chiu,M.W.,
Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
Gibbs,R.A.

JOURNAL
REFERENCE
AUTHORS

Unpublished
5 (bases 1 to 326663)
Chiu,M.W.

TITLE
JOURNAL

Direct Submission
Submitted (27-JUN-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

6 (bases 1 to 326663)
Gorrell,J.H.

TITLE
JOURNAL

Direct Submission
Submitted (17-NOV-1998) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
On Aug 20, 1997 this sequence version replaced gi:1710997

COMMENT

Sequencing is completed to a minimum standard of double strand
coverage or two chemistries on at least two separate clones, with
three overlapping reads. If sequence quality does not meet this
standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.

FEATURES
source

Location/Qualifiers
1. 326663
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xq28"
/clone="U126G1, U142F2, U69B6, U145C10, U169A5, U84H1,
U24D12, U80A7, U153E6, R7-163A8"

/clone_lib="LLNL X chromosome library LLOXNC01"
1. 1305
/rpt_family="L1"

repeat_region
1756..2531
/rpt_family="LIPAL1"

repeat_region
4587..5254
/rpt_family="LIMC2"

repeat_region
4599..4859
/rpt_family="LIPB3"

repeat_region
8069..8742
/rpt_family="LIME2"

repeat_region
complement(8846..9882)
/rpt_family="LIMA2"

repeat_region
complement(9883..10657)
/rpt_family="L1"

repeat_region
complement(10703..10991)
/rpt_family="Alu"

repeat_region
12190..12452
/rpt_family="MER43"

repeat_region
13055..13101
/rpt_family="MER43"

repeat_region
14402..14728
/rpt_family="MER28"

repeat_region
complement(17836..18198)
/rpt_family="THE1b"

repeat_region
complement(18199..19751)

```
/rpt_family="THR"
complement(19772..20114)
repeat_region
/rpt_family="THE1b"
26241..26504
repeat_region
/rpt_family="MER2"
27126..27410
repeat_region
/rpt_family="Alu"
complement(27721..28090)
repeat_region
/rpt_family="MLT1b"
complement(29141..29221)
repeat_region
/rpt_family="MIR"
complement(29971..30214)
STS
/db_xref=dbSTS:G04466"
complement(31323..32192)
repeat_region
/rpt_family="L1"
32202..33477
repeat_region
/rpt_family="L1"
33649..33828
repeat_region
/rpt_family="MLT1f"
complement(34208..34640)
STS
/db_xref=dbSTS:L18338"
complement(37847..37896)
repeat_region
/rpt_family="MIR"
complement(37897..38047)
repeat_region
/rpt_family="THE1b"
39003..39413
repeat_region
/rpt_family="Alu"
39447..39616
repeat_region
/rpt_family="Alu"
complement(42772..42890)
repeat_region
/rpt_family="MIR2"
43946..44051
repeat_region
/rpt_family="L1PA7"
44392..44469
repeat_region
/rpt_family="MIR2"
complement(44547..44844)
repeat_region
/rpt_family="L1"
complement(44932..45060)
repeat_region
/rpt_family="L1PA15"
complement(45326..45416)
repeat_region
/rpt_family="MIR2"
45428..45935
repeat_region
/rpt_family="MLT2B2"
47122..47237
repeat_region
/rpt_family="MIR"
complement(49106..49258)
repeat_region
/rpt_family="Alu"
complement(56772..56846)
repeat_region
/rpt_family="Alu"
57858..58215
repeat_region
/rpt_family="MLT1a"
complement(58322..58680)
repeat_region
/rpt_family="MLT1a"
complement(58693..58780)
repeat_region
/rpt_family="MLT1a"
58884..59162
repeat_region
/rpt_family="Alu"
complement(59176..59299)
repeat_region
/rpt_family="MStc"
complement(61342..61403)
repeat_region
/rpt_family="MIR"
63434..63486
repeat_region
/rpt_family="MIR"
complement(64072..64357)
repeat_region
/rpt_family="Alu"
64390..64546
repeat_region
/rpt_family="MIR"
complement(65193..65451)
STS
/db_xref=dbSTS:Z23490"
65883..66218
repeat_region
/rpt_family="MStA"
68209..68312
repeat_region
/rpt_family="MIR"

59326..69431
/rpt_family="MLT2C2"
complement(69641..69770)
STS
/db_xref=dbSTS:L39873"
73923..74072
repeat_region
/rpt_family="MER42c"
75581..75674
repeat_region
/rpt_family="MIR"
complement(77945..78229)
repeat_region
/rpt_family="LTR1"
complement(81874..81938)
repeat_region
/rpt_family="MIR"
82148..82576
repeat_region
/rpt_family="LTR7"
86925..86957
repeat_region
/rpt_family="LTR7"
87860..88147
repeat_region
/rpt_family="MER4B"
90287..90536
repeat_region
/rpt_family="MLT1e"
92315..92507
repeat_region
/rpt_family="MIR"
complement(93107..93413)
misc_feature
/note="similar to EST with GenBank Accession Number
Z33565"
repeat_region
93873..94223
/rpt_family="MLT1a"
repeat_region
complement(94579..94636)

...: remainder of annotations omitted.

Query Match 43.3%; Score 995; DB 9; Length 326663;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248953 GAAAAGAAACATGGATTACATGTTAAACATAATATTTTGAATATATATAGGCTCATAGAT 249012
Cp 2296 GAAAAGAAACATGGATTACATGTTAAACATAATATTTTGAATATATATAGGCTCATAGAT 2237
Db 249013 ATATTATTAAATATATAACAAATATATCAGGAACITAGTGTGGTGCCTAGTTTGATATAT 249072
Cp 2236 ATATTATTAAATATATAACAAATATATCAGGAACITAGTGTGGTGCCTAGTTTGATATAT 2177
Db 249073 GATTACTTTTGAATGCACTAAATCCACAAATATGAAAGTATTTCTTTGTGTAATG 249132
Cp 2176 GATTACTTTTGAATGCACTAAATCCACAAATATGAAAGTATTTCTTTGTGTAATG 2117
Db 249133 TTATGTTGGTTATTATGATGTTGCTCTTCGTATCCAAAGGTATGACATAACTTGAGTTTGT 249192
Cp 2116 TTATGTTGGTTATTATGATGTTGCTCTTCGTATCCAAAGGTATGACATAACTTGAGTTTGT 2057
Db 249193 TTGCTTTGTATTATTCAGTAATAAAGCCGTAACTGTTTAAAAAGAGGAAATTTAAAAA 249252
Cp 2056 TTGCTTTGTATTATTCAGTAATAAAGCCGTAACTGTTTAAAAAGAGGAAATTTAAAAA 1997
Db 249253 AAAAAGTGTCCAATTACCAATTATAAATTTTAATAAAGTAAACGAAAGGTTTGGCT 249312
Cp 1996 AAAAAGTGTCCAATTACCAATTATAAATTTTAATAAAGTAAACGAAAGGTTTGGCT 1937
Db 249313 GTTACATATTCTCAGGCCAAATTTGTTGATGTCATGTTTGGATTAACTAGCCCTCAGCTGC 249372
Cp 1936 GTTACATATTCTCAGGCCAAATTTGTTGATGTCATGTTTGGATTAACTAGCCCTCAGCTGC 1877
Db 249373 TTCCAATATTATGGGTAATFACAAAACGACCGCTAACTCCTCTCTCACCAGCTGGA 249432
Cp 1876 TTCCAATATTATGGGTAATFACAAAACGACCGCTAACTCCTCTCTCACCAGCTGGA 1817
Db 249433 AGGGAGCACATCACATTTGCCATCCATGTTGGCAAAAACCTCAAGGCATCAACACTGGAA 249492
Cp 1816 AGGGAGCACATCACATTTGCCATCCATGTTGGCAAAAACCTCAAGGCATCAACACTGGAA 1757
Db 249493 AAGATCTCCACCTTGGGAATCATTTATACATATTGTGATCCTGCAATGGGTGAGATCCAC 249552
|||||
```


BASE COUNT 494 a 375 c 444 g 518 t
ORIGIN

Query Match 18.9%; Score 434; DB 10; Length 1831;
Best Local Similarity 85.0%; Pred. No. 0.00e+00;
Matches 533; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

```
Db 1 CTTCTGGCAGTTGGGTACACACAGCCGCCATATCCCTTCAGATACCCAGGAATTTCA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 781 CTTCTGGCAGTTGGGTATAGCCACACATCCCTTCAGATACCCAGGAATTTCA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAAGTTGCATCCCTTGGAAACATAACCCCTGCTCTCATCCCATGTTCTCTGATAGCT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 841 GAAGTTGATCCCTTGGAAACATACCCCTGCTCTCATCCCATGTTCTCTGATAGCT 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 ACCACTGTAGCCCTACACCCCTGGATGGATATCAGGAGAGGGAAGATGTCACAGCCTT 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 901 ACCCCCTGTGSCCTACACCCCTGGATGGATATCAGGAGAGGGAAGATGTCACAGCCTT 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 AAACATCAGTGTCCCTATGAGCAAAATTCCTGAGGATTTTCAGCGGAAAATCCGCCAGAG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 961 AAACATCAGTGTCCCTATGAGTGTCCAAATTCCTGAGGATTTTCAGCGGAAAATCCGCCAGAG 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CTACTTTGCTCTGTGCTATTTGGATACATCAGTCAGTCGGACATGTTTGAAGTCTTGA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1021 CTACTTTGCTCTGTGCTATTTGGATACATCAGTCAGTCGGACATGTTTGAAGTCTTGA 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TGATCTTCGTCTAGCCACACAAATATTTGCTTTTCATCTGATCAGGATGGCCCT 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1081 CGATCTTCAGTGGCCACACACACCATATTTGATTTACCTCGGATCATGGTGGGCTCT 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AGGTGAACACGGAATGGGCAAGTACAGCAACTTTGATGTGCTACCCGTGTCGCCCT 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1141 AGGTGAACATGGAGATGGGCAAAATACAGCAATTTTGTGATGTGCTACCCATGTTCCCT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GATGCTATGTCCTTGAAGAGCGGCCCACTTCCTCGGAGCGCCAGAGCTCTTTCC 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1201 GATATTTATGTTCTGGAAGAGCGGCTTCACTTCCGAGGCGAGAGCTTTTCC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATACCGGAGCCCTTTGATCCGGCTCGGATTTGATGATCAGGACGACACCCAGGA 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1261 TTACTCGACCTTTTGAATCCGGCTACAGTTGATGAGCGCAGGCAATCCATGGA 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CCTCGTGAAGTCTATCTCTTCCCGACGCTCGCTGGACTTGCGAGACTGCCAGT-CC 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1321 CCTGTGGAAGTCTGCTCTCTTTTCCACGCTGGCTGGCTGGAGACTTGCGAGACTGCCAGTCC 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 TCCTCGGTGCCCCATCCCTTTTCA 626
Qy 1381 ACCTCGCTGCCCGTTCCTTCATTTCA 1407
```

RESULT 13
LOCUS G13552 400 bp DNA STS 04-JAN-1996
DEFINITION human STS SHGC-11149.
ACCESSION G13552
VERSION G13552.1 GI:1129291
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choenata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catarrhini; Hominoidea; Homo.
1 (bases 1 to 400)
Myers, R.M.
Unpublished (1995)

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687

Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: TGGTGAGGAGGAGGAGTTAGAC
Primer B: GGCTGTTACATATTTCTCAGGCC
STS size: 118
PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3

Prepared with primer pairs derived from M58342 -- Unigene.

FEATURES
source
STS
primer_bind
primer_bind
BASE COUNT_132 a 69 c 68 g 121 t 10 others
ORIGIN

Query Match 17.0%; Score 390; DB 11; Length 400;
Best Local Similarity 97.5%; Pred. No. 1.49e-290;
Matches 390; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Db 1 GTTTGGCCCAACCATGGATGGCAAAATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAGG 60
Qy 1778 GTTTGGCCCAACCATGGATGGCAAAATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAGG 1837
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TTAGAGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCCTGAGGCTAGTTAAT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1838 TTAGAGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCCTGAGGCTAGTTAAT 1897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CCAAAACATGCATCAACAATTTGGCCCTGAGAATATGTAACAGCCAAACCTTTTCGTTTAGT 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1898 CCAAAACATGCATCAACAATTTGGCCCTGAGAATATGTAACAGCCAAACCTTTTCGTTTAGT 1957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CTTTATTAATAATTTAATTTGTTAAATTTGGACCAAGNNNNNNNNNNNAATTTCCCTCTTTT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1958 CTTTATTAATAATTTAATTTGTTAAATTTGGACCAAGTTTTCCTCTCTTTT 2017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AAACAGTTACGGCTTATTTACTGATAATAATACAAACCAACAACTCAAGTTATGTCATA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2018 AAACAGTTACGGCTTATTTACTGATAATAATAACAAACCAACAACTCAAGTTATGTCATA 2077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 CTTTGGATACCAAGCACCATACATATATACCAACATACATATATACACAAAGAACTT 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2078 CTTTGGATACCAAGCACCATACATATATACCAACATACATATATACACAAAGAACTT 2137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TCATTATTTGTGGAATTTAGTGCATTTCAAAAAGTAATCA 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2138 TCATTATTTGTGGAATTTAGTGCATTTCAAAAAGTAATCA 2177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 14
LOCUS AC002315 110079 bp DNA ROD 20-FEB-1998
DEFINITION Mouse BAC-146N21 Chromosome x contains iduronate-2-sulfatase gene;

complete sequence.

AC002315

AC002315.1 GI:2258164

HTG

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 110079)

Timms,K.M., Hockett,L., Edwards,F.J., Lu,J., Muzny,D.M., Miller,W.

and Gibbs,R.A.

TITLE

Cross-species sequence comparison of the IDS region

JOURNAL

REFERENCE

2 (bases 1 to 110079)

Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,

Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,

Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,

Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,

Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and

Gibbs,R.A.

TITLE

Direct Submission

JOURNAL

REFERENCE

3 (bases 1 to 110079)

Chiu,M.W.

TITLE

Direct Submission

JOURNAL

Submitted (15-JUL-1997) Molecular and Human Genetics, Baylor

College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Sequencing is completed to a minimum standard of double strand

coverage with a minimum of 2 clones and 2 reads with no ambiguities

or 2 chemistries with a minimum of 2 clones and 3 reads with no

ambiguities. If the sequence quality does not meet this standard,

it will be indicated in the annotation.

COMMENT

FEATURES

source

1. .110079

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="X"

/clone="BAC-146N21"

244..561

/rpt_family="LTR/MaLR"

567..2305

/rpt_family="LTR/MaLR"

2306..2646

/rpt_family="LTR/MaLR"

complement(2864..3101)

/rpt_family="LTR/Retroviral"

complement(3812..3847)

/rpt_family="Simple_repeat"

complement(4982..5064)

/rpt_family="Simple_repeat"

complement(5065..5184)

/rpt_family="LINE/L1"

complement(5257..5457)

/rpt_family="SINE/B2"

6904..6953

/rpt_family="Simple_repeat"

complement(7980..8199)

/rpt_family="SINE/B2"

complement(10494..10582)

/rpt_family="SINE/Alu"

complement(10633..10861)

/rpt_family="LTR/MaLR"

complement(10892..11185)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/rpt_family="SINE/B4"

complement(11193..11366)

/rpt_family="LTR/MaLR"

11629..11718

/rpt_family="Simple_repeat"

complement(11839..11971)

/rpt_family="SINE/B4"

complement(12138..12171)

/rpt_family="Low_complexity"

complement(12276..12396)

/rpt_family="SINE/B2"

14061..14118

/rpt_family="Simple_repeat"

14096..14220

/note="Single strand coverage only"

14179..14201

/rpt_family="Low_complexity"

join(14201..14287,15145..15281,15929..16106,17952..18040,

18934..19134,20646..20816,25808..25934,27982..28155)

/gene="IDS"

complement(14424..14541)

/rpt_family="Simple_repeat"

complement(14783..14914)

/rpt_family="Simple_repeat"

17009..17084

/rpt_family="SINE/MIR"

complement(17555..17697)

/rpt_family="SINE/Alu"

complement(17620..17719)

/rpt_family="LTR/MaLR"

complement(17821..17869)

/rpt_family="SINE/MIR"

complement(18187..18314)

/rpt_family="SINE/Alu"

18371..18499

/rpt_family="SINE/B4"

19306..19411

/rpt_family="SINE/Alu"

complement(19902..20388)

/rpt_family="Other/MER21_gro"

complement(21582..21661)

/rpt_family="Simple_repeat"

complement(21662..21761)

/rpt_family="scRNA"

complement(21684..21761)

/rpt_family="SINE/ID"

21812..21849

/rpt_family="DNA/MER1_type"

complement(23849..23940)

/rpt_family="SINE/Alu"

25152..25220

/rpt_family="LTR/MaLR"

complement(26165..26376)

/rpt_family="SINE/B2"

complement(26379..26577)

/rpt_family="SINE/B2"

26995..27125

/rpt_family="SINE/B2"

27322..27505

/rpt_family="SINE/B2"

complement(27695..27734)

/rpt_family="Low_complexity"

29292..30131

/rpt_family="LINE/L1"

30204..30233

/rpt_family="Simple_repeat"

complement(31179..31235)

/rpt_family="Simple_repeat"

31236..31266

/rpt_family="Simple_repeat"

31579..31703

/rpt_family="Simple_repeat"

31705..32083

W P S R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:16:13 2000; MasPar time 44.92 Seconds
Tabular output not generated. 848.838 Million cell updates/sec

Title: >US-09-249-003-2
Description: (1-550) from US09249003.pep
Perfect Score: 4069
Sequence: 1 MPPTGRLGLLGLVLSV.....QDHNMYNDGGDLFQLLMP 550

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_plant 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 50.386; Variance 92.586; scale 0.544

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2445	60.1	343	4	IDURONATE-2-SULPHATASE	0.00e+00
2	1036	25.5	179	4	IDURONATE-2-SULPHATASE	1.00e-194
3	604	14.8	86	4	IDURONATE-2-SULPHATASE	8.45e-99
4	338	8.3	512	2	CHOLINE SULFATASE (EC	1.46e-42
5	190	4.7	551	5	ARYLSULFATASE	3.92e-14
6	185	4.5	514	2	PHOSPHONATE MONOESTER	2.97e-13
7	182	4.5	970	2	HYPOTHETICAL 105.7 KD	9.93e-13
8	179	4.4	465	2	ARYSULFATASE	3.29e-12
9	174	4.3	559	5	ARYLSULFATASE	2.39e-11
10	143	3.5	452	5	ARYLSULFATASE	3.09e-06
11	134	3.3	525	4	KIAA1001 PROTEIN.	7.70e-05
12	136	3.3	787	2	ARYLSULFATASE	3.80e-05
13	133	3.3	787	2	ARYLSULFATASE	1.09e-04
14	126	3.1	571	2	HYPOTHETICAL 86.1 KD P	1.22e-03
15	123	3.0	577	2	ARYLSULFATASE PRECURSOR	3.35e-03
16	121	3.0	580	2	POLAR FLAGELLAR M-RING	6.51e-03
17	117	2.9	609	3	HYPOTHETICAL C2H2-TYPE	2.41e-02
18	112	2.8	495	2	PUTATIVE 54.5 KDA PROT	1.19e-01
19	108	2.7	371	8	CYTOCHROME OXIDASE I (4.12e-01
20	108	2.7	374	2	OMP37 PORIN PRECURSOR	4.12e-01

21	111	2.7	409	8	Q9XP15	CYTOCHROME C OXIDASE S	1.63e-01
22	111	2.7	708	1	Q9YC40	708AA LONG HYPOTHETICA	1.63e-01
23	109	2.7	709	5	Q21376	SIMILAR TO SULFATASE.	3.03e-01
24	108	2.7	1811	5	Q9X2U8	LARGEST SUBUNIT OF THE	4.12e-01
25	109	2.7	1889	5	Q9X2U9	LARGEST SUBUNIT OF THE	3.03e-01
26	109	2.7	1889	5	Q9X2S2	LARGEST SUBUNIT OF THE	3.03e-01
27	108	2.7	1889	5	Q9X2S1	LARGEST SUBUNIT OF THE	4.12e-01
28	111	2.7	7576	2	Q9ZGA4	FK506 POLYTIDE SYNTH	1.63e-01
29	104	2.6	159	2	Q9Z3T1	HALOALKANOIC ACID DEHA	1.38e+00
30	105	2.6	403	8	Q9XP03	CYTOCHROME C OXIDASE S	1.02e+00
31	107	2.6	409	8	Q9XP02	CYTOCHROME C OXIDASE S	5.59e-01
32	107	2.6	410	8	Q9XP10	CYTOCHROME C OXIDASE S	5.59e-01
33	105	2.6	647	2	O84272	TRANSGLYCOLASE/TRANSP	1.02e+00
34	106	2.6	667	2	P71538	PROBABLE CARBAMOYL-PHO	7.58e-01
35	106	2.6	815	5	Q18385	SIMILAR TO PROTEIN TYR	7.58e-01
36	107	2.6	1034	11	Q61856	ORF.	5.59e-01
37	107	2.6	3262	4	Q13788	APOLIPOPROTEIN B-100 (5.59e-01
38	101	2.5	280	2	P71422	MDCB.	3.34e+00
39	102	2.5	366	14	O86317	NON-STRUCTURAL SIGMA-N	2.49e+00
40	101	2.5	366	14	O86308	NON-STRUCTURAL SIGMA-N	3.34e+00
41	103	2.5	409	8	Q9XP12	CYTOCHROME C OXIDASE S	1.86e+00
42	103	2.5	410	8	Q9XP16	CYTOCHROME C OXIDASE S	1.86e+00
43	103	2.5	440	2	P75608	R02_ORF440 PROTEIN.	1.86e+00
44	103	2.5	467	2	O06256	HYPOTHETICAL 47.6 KD P	1.86e+00
45	103	2.5	518	2	O68838	GAMMA-GLUTAMYL-CYSTEINE	1.86e+00

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	343 AA.
ID	Q14604			
AC	Q14604			
DT	01-NOV-1996 (TReMBLrel. 01, Created)			
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)			
DE	IDURONATE-2-SULPHATASE.			
GN	IDS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LYMPHOCTE;			
RX	MEDLINE; 96079126.			
RA	MALMGREN H., CARLBERG B.M., PETERSSON U., BONDESON M.L.;			
RT	"Identification of an alternative transcript from the human			
RT	iduronate-2-sulfatase (IDS) gene."			
RL	Genomics 29:291-293(1995).			
DR	EMBL; L40586; AAA92014.1; -			
DR	PROSITE; PS00149; SULFATASE_2; 1.			
DR	PROSITE; PS00523; SULFATASE_1; 1.			
DR	PFAM; PF00884; Sulfatase; 1.			
SQ	SEQUENCE 343 AA; 38310 MW; F1825B26 CRC32;			

Query Match 60.1%; Score 2445; DB 4; Length 343;
Best Local Similarity 99.1%; Pred. No. 0.00e+00;
Matches 336; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db	1	MPPTGRLGLLGLVLSVVCVALGSETQANSTTDALNVLLIIVDDLPSICGCGDKLVR	60
QY	1	MPPTGRLGLLGLVLSVVCVALGSETQANSTTDALNVLLIIVDDLPSICGCGDKLVR	60
Db	61	SPNIDQLASHLSLFFQNAFAQAQVACAPSRVSLTGRRRDPTTLRLYDFNSYWRVHAGNFSTIP	120
QY	61	SPNIDQLASHLSLFFQNAFAQAQVACAPSRVSLTGRRRDPTTLRLYDFNSYWRVHAGNFSTIP	120
Db	121	QYFKENGYVTMSYGVKVFHPGPGISSNHTDDSPYSFSPFHPSPSEKYEKTKTCRPGDGLHA	180
QY	121	QYFKENGYVTMSYGVKVFHPGPGISSNHTDDSPYSFSPFHPSPSEKYEKTKTCRPGDGLHA	180
Db	181	NLLCPVDVLDVPGTLPDKOSTEQATQLLEKMKTSASPFPLAVGYHKPHTPFPYKPEFOK	240
QY	181	NLLCPVDVLDVPGTLPDKOSTEQATQLLEKMKTSASPFPLAVGYHKPHTPFPYKPEFOK	240


```

RESULT 5
ID Q25384 PRELIMINARY; PRT; 551 AA.
AC Q25384;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE ARYL SULFATASE.
GN ARS.
OS Hemicentrotus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Hemicentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RA YAMADA K.;
RL Submitted (SEP-1989) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 1-105 FROM N.A.
RP TISSUE=SPERM;
RC MEDLINE; 90092130.
RA YAMADA K., AKASAKA K., SHIMADA H.;
RX "Structure of sea-urchin arylsulfatase gene.";
RT Eur. J. Biochem. 186:405-410(1989).
[3]
RN SEQUENCE FROM N.A.
RP TISSUE=SPERM;
RC SHIMADA H.;
RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP TISSUE=SPERM;
RC AKASAKA K., SAKAMOTO N., YAMAMOTO T., MOROKUMA J., FUJIKAWA N.,
RA TAKATA K., EGUCHI S., SHIMADA H.;
RL Dev. Growth Differ. 36:633-636(1994).
DR EMBL; X15679; CAA34667.1; -.
DR EMBL; X15680; CAA34667.1; JOINED.
DR EMBL; X15681; CAA34667.1; JOINED.
DR EMBL; X15682; CAA34667.1; JOINED.
DR EMBL; X15683; CAA34667.1; JOINED.
DR EMBL; X15684; CAA34667.1; JOINED.
DR HSP; P15289; 1AUK.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PRAM; PF00884; Sulfatase; 1.
SQ SEQUENCE 551 AA; 60943 MW; 58CE54B7 CRC32;

Query Match 4.7%; Score 190; DB 5; Length 551;
Best Local Similarity 33.6%; Pred. No. 3.92e-14;
Matches 40; Conservative 28; Mismatches 44; Indels 7; Gaps 6;

Db 53 NVLLVADDDGSGDLTSYGHPTQAGIDKMAEGLRFTNGYVGDAVCTPSRSALMTGRL 112
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 38 NVLLIIVDDLRLPS-LGCGDKLVRSPNIDQLASHSLLFQNAFAQAQCAPSRVSFLTGR 96
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 113 PVRIGTGETRVLFWTKTLGPKRSELTIAEAMKEAGYATCMVK-WHLGINENSSIDGA 170
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 97 PDTRLX-DFNSY--W-RVHAGNPS-TIOPYFKENGYYTMSVGKVFHPGSISSNHTDDSP 150
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 6
ID Q45087 PRELIMINARY; PRT; 514 AA.
AC Q45087;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE PHOSPHONATE MONOESTER HYDROLASE.
GN PEHA.
OS Burkholderia caryophylli.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
RN [1]
RP SEQUENCE FROM N.A.

```

```
Db 268 GRNHRVGVGSCBFPYPCYSASVPSCAALPRILRDNCGVYTCAGCK-WHLTPDNYQG 326
QY 94 GR---RDTTLYDF-NSY---WR-VHAGNFSTIPQYFKENGIVTMSVGKVFHPGSSNHT 146
Db 327 AAGPDFNWPL 336
QY 147 DDSPY-SWSF 155

RESULT 8
ID O53700 PRELIMINARY; PRT; 465 AA.
AC O53700; O07223;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE ATSG PROTEIN.
GN ATSG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY III C.E., TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDRAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA PARKHILL J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL021930; CAAL1731.1; -.
DR PFAM; PF00884; Sulfatase; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51846 MW; D65FE5FA CRC32;

Query Match 4.4%; Score 179; DB 2; Length 465;
Best Local Similarity 35.2%; Pred. No. 3.29e-12;
Matches 37; Conservative 25; Mismatches 39; Indels 4; Gaps 4;

Db 3 SERATQRENLLVHHDGLGRYLGVYHHPDVYSPRLDRLAAEGILEFTRAHATAPLCTPSR 62
QY 30 ANSTTDALNLLIIV-DDLRPSLCYGDKLVRSFNIDQLASHLLFQNAFAQAVCAPSR 88
Db 63 GSFLTGTGYPQSGNLGLAHGHWERTGY-QTLPLLSESGWYSAL 106
QY 89 VSFLTGRPDTRLYDFNSY-WRVHAGNFSTIPQYFKENG-YVTM 131

RESULT 9
ID O16138 PRELIMINARY; PRT; 559 AA.
AC O16138;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ARYLSULFATASE.
GN ARS.
OS Helicoidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoidea; Echinometridae; Helicoidaris.
RN [1]
RP SEQUENCE FROM N.A.
RA HAAG E.S., RAFF R.A.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; AF013158; AAC27821.1; -.
DR HSSP; P15289; IAUk.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PFAM; PF00884; Sulfatase; 1.
SQ SEQUENCE 559 AA; 61915 MW; 731B880D CRC32;

Query Match 4.3%; Score 174; DB 5; Length 559;
Best Local Similarity 30.8%; Pred. No. 2.39e-11;
Matches 37; Conservative 31; Mismatches 43; Indels 9; Gaps 8;

Db 59 NVILMVADDMGVGDLSVYCHP-TQEPGFIDEMAANGCLRTNGYVDSVCTPSPSAIMTGR 117
QY 38 NVLLIIVDDL-RPSLCYGDKLVRSFN-IDQLASHLLFQNAFAQAVCAPSRVSLTGR 95
Db 118 LPRIQTGEIRVFLPWTKTGLPKSEVTIAEQMGQGYRTGMVGK-WHLIGNEQTSTDGA 176
QY 96 RPDTRLY-DFNSY--W-RVHAGNFS-TIPQYFKENGIVTMSVGKVFHPGSSNHTDDSP 150

RESULT 10
ID Q18924 PRELIMINARY; PRT; 452 AA.
AC Q18924;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE SIMILAR TO ARYLSULFATASE.
GN D1014.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA DU Z., LEIMBAC D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53180; AAA96290.1; -.
DR HSSP; P15289; IAUk.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PFAM; PF00884; Sulfatase; 1.
SQ SEQUENCE 452 AA; 51685 MW; FDAC4E64 CRC32;

Query Match 3.5%; Score 143; DB 5; Length 452;
Best Local Similarity 31.1%; Pred. No. 3.09e-06;
Matches 19; Conservative 21; Mismatches 20; Indels 1; Gaps 1;

Db 34 NVILMIDDLGYGDIASGYCHPTQYVTRFQYVTRFQYVTRFQYVTRFQYVTRFQYVTR 93
QY 38 NVLLIIVDDL-RPSLCYGDKLVRSFNIDQLASHLLFQNAFAQAVCAPSRVSLTGR 96
```


RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).

DR EMBL; 295972; CAB09383.1; -
DR PFAM; PF00884; Sulfatase; 1.
KW Hypothetical protein.
SQ SEQUENCE 787 AA; 86127 MW; 5E38P5FF CRC32;

Query Match 3.3%; Score 133; DB 2; Length 787;

Best Local Similarity 32.9%; Pred. No. 1.09e-04; Mismatches 22; Conservative 24; Indels 3; Gaps 3;

Db 35 QLRPAGAPNVLLILLDDVGGSSAFGGP-CRTSTAEALLAGNLRY-NREHTTALCSPT 92

QY 29 QANSTTDALNVLLIIVDDL-R-PSLGCYGDKLVRSPNIDQLASHLLFQNAFAQAVCAPS 87

Db 93 QOALLTCGRNHSA 105

QY 88 RVSELTGRRPDTT 100

RESULT 14

ID P77318 PRELIMINARY; PRT; 571 AA.

AC P77318; P78159;

DT 01-FEB-1997 (TremBLrel. 02, Created)

DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)

DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)

DE HYPOTHETICAL 64.1 KD PROTEIN.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN=K-12;

RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RC SEQUENCE FROM N.A.

RP STRAIN=K12;

RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,

RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,

RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,

RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,

RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,

RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN [3]

RC SEQUENCE OF 90-571 FROM N.A.

RP STRAIN=K12;

RX MEDLINE; 97251357.

RA AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K., ITOH T.,

RA KASAI H., KASHIMOTO K., KIMURA S., KITAGAWA M., KITAGAWA M.,

RA MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,

RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N.,

RA SAMPEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H., TAKEADA J.,

RA TAKEMOTO K., TAKEUCHI Y., WADA C., YAMAMOTO Y., HORIUCHI T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 28.0-40.1 min region on the linkage map.;"

RL DNA Res. 3:363-377(1996).

DR EMBL; AE000247; AAC74571.1; -

DR EMBL; D90792; BAA15172.1; -

DR EMBL; D90791; BAA15169.1; -

DR PFAM; PF00884; Sulfatase; 1.

KW Hydrolase.

SQ SEQUENCE 571 AA; 64094 MW; 447178C5 CRC32;

Query Match 3.1%; Score 126; DB 2; Length 571;

Best Local Similarity 25.0%; Pred. No. 1.22e-03;

Mismatches 49; Conservative 52; Mismatches 78; Indels 17; Gaps 16;

Db 120 TPTLLSMDGVRFTNGYVAHGVSGPSRAALMTGRAPARFQVYNTDAQDGIPLTE-TFL 178

QY 61 SPNIDQLASHLLFQNAFAQAVCAPSRVSLTGRRPDTTRLY-DFNSYWRVHAGNFSTI 119

Db 179 PELFQNHGYTAAVGK-WHLISKISNPVPEDKQTRDYHDNETTFSAEWQPNRGFDYFM 237

QY 120 POYFKENGIVTMSGVKVFHPG-ISSNHTDDSPYSWSFPP-YHP-SSEKYE-NTKT--C-R 172

Db 238 GFHAAGTAYNSPSLFRKRVPRKGYISD-QLTDEAIGVVDRAKTLDOFFMLYLAYNAP 296

QY 173 GPDGELHANLLCP-V-DVLD-VPE-GTLPDKOSTEQAIQLEKMKTSASPFPLAVGYHKP 228

Db 297 HLPNDNPADQYQKQF 312

QY 229 HIPFRYPK-E-FOKLY 242

RESULT 15

ID Q9X759 PRELIMINARY; PRT; 577 AA.

AC Q9X759;

DT 01-NOV-1999 (TremBLrel. 12, Created)

DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)

DE ARYLSULFATASE PRECURSOR (EC 3.1.6.1).

GN ATSA.

OS Klebsiella pneumoniae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Klebsiella.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN=DSM 681;

RX MEDLINE; 99269066.

RA SZAMET C., MIECH C., BALLEININGER M., SCHMIDT B., VON FIGURA K.,

RA DIERKS T.;

RT "The iron sulfur protein AtsB is required for posttranslational

formation of formylglycine in the Klebsiella sulfatase.;"

RL J. Biol. Chem. 274:15375-15381(1999).

DR EMBL; AJ131525; CAB40961.1; -

DR PROSITE; PS00149; SULFATASE_2; 1.

DR PROSITE; PS00523; SULFATASE_1; 1.

KW Signal; Hydrolase.

FT SIGNAL

SQ SEQUENCE 577 AA; 64155 MW; C1ECC635 CRC32;

Query Match 3.0%; Score 123; DB 2; Length 577;

Best Local Similarity 18.8%; Pred. No. 3.35e-03;

Mismatches 63; Conservative 107; Mismatches 138; Indels 27; Gaps 21;

Db 8 AAVSMILAGGAHA-AQOERPNIIVIIADDMGYSIDTSPFGGE-IPTPNLCMAEQGRMSQ 65

QY 18 SSVCVALGSETQANSTTDALNVLLIIVDDL-R-PSLGCYGDKLVRSPNIDQLASHLLFON 76

Db 66 YTS-PMSAPARSMILLTGNSSQAGMGWYDSTIGKEGYELRL-TDRVTTMAERFKA 123

QY 77 AFAQAVCAPSRVSLTGRR-PDTTR--L-YD--F--NSY-WRVHAGNFSTIPQYFKN 126

Db 124 GYNTLMAGK-WHLGVPVGPATPKERGFNHAFAFMGGTSHFNDAIP-LGTVEAFHYTYTRD 181

QY 127 GYVTSVGVKVFHPG-ISSNHTDDSPYSWSFPPHPSPSEKYENTKTCRGDGLHANLLCP 185

Db 182 GERSVLPDDFVSSEAYARQMNSWI-KATPKQPVPAWLAFTAPHDPLQAPDEWIKRFQKQ 240

QY 186 VDVLVPEGTLPDKOSTEQAIQLEKMKTSASPFPLAVGYHHRPHIPFRYPKPFQKLYPLE 245

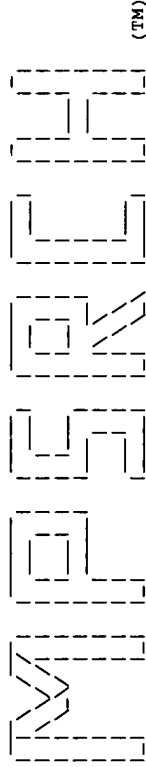
Db 241 YEQGYAEV-YRORIARL--KA-LGIHDDTLPPLHELDKEWALTPQCKYKAKYMOVVA 296

QY 246 NITLAPDPEVDPGLPPVAYNPMWDIRQRED-VQALNISVPYGPV-PVD--FORKINQSYF 301

Db 297 AMIANMDAQIGTMMETLKTQGRDKNTLLVFLTDNG 331

QY 302 ASVSYLDTQVGRLLSALDDQLANSTIIFTSDHG 336

Search completed: Wed Apr 19 21:34:11 2000
Job time : 1078 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mprsch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:09:19 2000; MasPar time 19.58 Seconds
Tabular output not generated. 838.906 Million cell updates/sec

Title: >US-09-249-003-2
Description: (1-550) from US09249003.pap
Perfect Score: 4069
Sequence: 1 MPPPTGRGLLWGLVLSV.....QDHNNYNDSSGGDLFQLLMP 550

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 52.053; Variance 88.452; scale 0.588

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4069	100.0	550	1	IDS_HUMAN	0.00e+00
2	3129	76.9	563	1	IDURONATE 2-SULFATASE	0.00e+00
3	282	6.9	497	1	YIDI_ECOLI	6.79e-34
4	266	6.5	624	1	STSL_MOUSE	1.50e-30
5	258	6.3	591	1	ARSL_MOUSE	6.75e-29
6	239	5.9	589	1	ARSL_HUMAN	5.12e-25
7	234	5.8	593	1	ARSL_HUMAN	5.22e-24
8	211	5.2	522	1	GA6S_HUMAN	1.89e-19
9	211	5.2	567	1	ARSL_STRPU	1.89e-19
10	213	5.2	583	1	STSL_HUMAN	7.70e-20
11	202	5.0	577	1	STSL_MOUSE	1.05e-17
12	198	4.9	506	1	ARSL_MOUSE	6.17e-17
13	195	4.8	507	1	ARSL_HUMAN	2.04e-15
14	190	4.7	533	1	ARSL_HUMAN	2.04e-15
15	184	4.5	551	1	ARSL_HUMAN	2.73e-14
16	176	4.3	502	1	SPHL_HUMAN	8.23e-13
17	172	4.2	535	1	ARSL_FELCA	4.42e-12
18	153	3.8	313	1	RIBF_ECOLI	1.04e-08
19	153	3.8	551	1	ARSL_ECOLI	1.04e-08
20	126	3.1	433	1	TCOL_HUMAN	2.92e-04
21	126	3.1	919	1	YK05_YEAST	2.92e-04
22	124	3.0	464	1	ARSL_KLEAE	5.95e-04
23	118	2.9	189	1	COAT_CCMV	4.83e-03

24	117	2.9	532	1	ARS_PSEAE	ARYLSULFATASE (EC 3.1.1.	6.80e-03
25	118	2.9	546	1	GCVK_HSV7J	POSSIBLE GANCICLOVIR K	4.83e-03
26	115	2.8	473	1	ARSL_MOUSE	ARYLSULFATASE B (EC 3.	1.34e-02
27	114	2.8	540	1	NADB_ECOLI	L-ASPARTATE OXIDASE (E	1.87e-02
28	114	2.8	559	1	GL6S_CAPHI	N-ACETYLGLUCOSAMINE-6-	1.87e-02
29	112	2.8	585	1	YEJM_HAEIN	HYPOTHETICAL PROTEIN H	3.65e-02
30	110	2.7	846	1	YDSB_SCHPO	HYPOTHETICAL 93.2 KD T	7.03e-02
31	105	2.6	285	1	ARSL_MOUSE	ARYLSULFATASE B (EC 3.	3.49e-01
32	107	2.6	552	1	GL6S_HUMAN	N-ACETYLGLUCOSAMINE-6-	1.85e-01
33	107	2.6	4563	1	APB_HUMAN	APOLIPOPROTEIN B-100 P	1.85e-01
34	101	2.5	190	1	YEJM_SALTY	HYPOTHETICAL PROTEIN I	1.20e+00
35	101	2.5	353	1	REBU_SALTY	REBU PROTEIN.	1.20e+00
36	101	2.5	383	1	OMS2_SALTY	OUTER MEMBRANE PROTEIN	1.20e+00
37	101	2.5	394	1	OMSL_SALTY	OUTER MEMBRANE PROTEIN	1.20e+00
38	101	2.5	510	1	MGSL_PSESM	2,3-BISPHOSPHOGlycerat	1.20e+00
39	102	2.5	518	1	GSHI_ECOLI	GLUTAMATE--CYSTEINE LI	8.85e-01
40	102	2.5	586	1	YEJM_ECOLI	HYPOTHETICAL 67.3 KD P	8.85e-01
41	100	2.5	1013	1	EPAS_CHICK	EPHRIIN TYPE-A RECEPTOR	1.62e+00
42	100	2.5	1896	1	RPBL_DROME	DNA-DIRECTED RNA POLYM	1.62e+00
43	98	2.4	366	1	RRPO_REOVD	RNA-DIRECTED RNA POLYM	2.95e+00
44	99	2.4	377	1	OMP_N_ECOLI	OUTER MEMBRANE PROTEIN	2.19e+00
45	99	2.4	2103	1	RRPL_UUK	RNA POLYMERASE (EC 2.7	2.19e+00

ALIGNMENTS

RESULT	ID	IDS_HUMAN	STANDARD;	PRT;	550 AA.
AC	P22304;				
DT	01-AUG-1991 (Rel. 19, Created)				
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).				
GN	IDS.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
RN	[1]				
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 34-58 AND 456-473				
RX	TISSUE-ENDOTHELIAL CELLS;				
RX	MEDLINE: 91046030.				
RA	WILSON P.J., MORRIS C.P., ANSON D.S., OCCHIODORO T., BIELICKI J.,				
RA	CLEMENTS P.R., HOPWOOD J.J.;				
RT	"Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone				
RT	and analysis of patient DNA.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:8531-8535(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 94063929.				
RA	WILSON P.J., MEANEY C.A., HOPWOOD J.J., MORRIS C.P.;				
RL	"Sequence of the human iduronate 2-sulfatase (IDS) gene.";				
RN	Genomics 17:773-775(1993).				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 96352905.				
RA	TIMMS K.M., LU F., SHEN Y., PIERSON C.A., MUZYNY D.M., GU Y.,				
RA	NELSON D.L., GIBBS R.A.;				
RT	"130 kb of DNA sequence reveals two new genes and a regional				
RT	duplication distal to the human iduronate-2-sulfate sulfatase				
RT	locus.";				
RL	Genome Res. 5:71-78(1995).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RA	MUZYNY D., ANSARI-LARI M.A., TIMMS K.M., YU W., DUGAN S., LU J.,				
RA	SHEN Y., ROWLAND K., LIU W., PEREZ L., DING Y., GONZALEZ O.,				
RA	HAYWOOD M., JAIN A., LEAL B., LOGAN O., NGUYEN V., SAVAGE L.,				
RA	WORLEY K., CHEN E., FORCUM J., ARENSON A.D., CHIU M.W., SHEN H.,				
RA	BRUNDAGE E., DI W., CHINAULT C., NELSON D., GIBBS R.A.;				
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.				
RN	[5]				
RP	SEQUENCE OF 1-398 FROM N.A.				
RX	MEDLINE: 93258349.				
RA	FLOMEN R.H., GREEN E.P., GREEN P.M., BENTLEY D.R., GIANNELLI F.;				

Applicant

RT "determination of the organisation of coding sequences within the
 RT iduronate sulphate sulphatase (IDS) gene.";
 RL Hum. Mol. Genet. 2:3-10(1993).
 [6]
 RN REVIEW ON MPS-II VARIANTS.
 RP MEDLINE; 94154729.
 RX HOPWOOD J.J., BUNGE S., MORRIS C.P., WILSON P.J., STEGLICH C.,
 RA BECK M., SCHWINGER E., GAL A.;
 RA "Molecular basis of mucopolysaccharidosis type II: mutations in the
 RT iduronate-2-sulphatase gene.";
 RL Hum. Mutat. 2:435-442(1993).
 [7]
 RN VARIANTS MPS-II ARG-135 AND GLY-422.
 RP MEDLINE; 9325059.
 RX BUNGE S., STEGLICH C., BECK M., ROSENKRANZ W., SCHWINGER E.,
 RA HOPWOOD J.J., GAL A.;
 RA "Mutation analysis of the iduronate-2-sulfatase gene in patients with
 RT mucopolysaccharidosis type II (Hunter syndrome).";
 RL Hum. Mol. Genet. 1:335-339(1992).
 [8]
 RN VARIANT MPS-II TRP-468.
 RP MEDLINE; 93258320.
 RX CROTTI P.L., BUNGE S., ANDERSON R.A., WHITLEY C.B.;
 RA "Mutation R468W of the iduronate-2-sulfatase gene in mild Hunter
 RT syndrome (mucopolysaccharidosis type II) confirmed by in vitro
 RL mutagenesis and expression.";
 RL Hum. Mol. Genet. 1:755-757(1992).
 [9]
 RN VARIANTS MPS-II ARG-86; ASP-94; ARG-120; PRO-221 AND GLY-422.
 RP MEDLINE; 94108441.
 RX BUNGE S., STEGLICH C., ZUTHER C., BECK M., MORRIS C.P., SCHWINGER E.,
 RA SCHINZEL A., HOPWOOD J.J., GAL A.;
 RA "Iduronate-2-sulfatase gene mutations in 16 patients with
 RT mucopolysaccharidosis type II (Hunter syndrome).";
 RL Hum. Mol. Genet. 2:1871-1875(1993).
 [10]
 RN VARIANTS MPS-II GLU-68; HIS-293; GLY-478 AND ARG-485.
 RP MEDLINE; 95072615.
 RX SCHROEDER W., WULF K., WEHNERT M., SEIDLITZ G., HERRMANN F.H.;
 RA "Mutations of the iduronate-2-sulfatase (IDS) gene in patients with
 RT Hunter syndrome (mucopolysaccharidosis II).";
 RL Hum. Mutat. 4:128-131(1994).
 [11]
 RN VARIANTS MPS-II TRP-132; TYR-229; ARG-358; HIS-469 AND CYS-523.
 RP MEDLINE; 95193786.
 RX JONSSON J.J., ARONOVICH E.L., BRAUN S.E., WHITLEY C.B.;
 RA "Molecular diagnosis of mucopolysaccharidosis type II (Hunter
 RT syndrome) by automated sequencing and computer-assisted
 RL interpretation: toward mutation mapping of the iduronate-2-sulfatase
 gene.";
 RL Am. J. Hum. Genet. 56:597-607(1995).
 [12]
 RN VARIANTS MPS-II LEU-86; ASN-87; PRO-92; ASN-135; LYS-345 AND TRP-468.
 RP MEDLINE; 95245347.
 RX POPOWSKA E., RATHMANN M., TYLKI-SZYMANSKA A., BUNGE S., STEGLICH C.,
 RA SCHWINGER E., GAL A.;
 RA "Mutations of the iduronate-2-sulfatase gene in 12 Polish patients
 RT with mucopolysaccharidosis type II (Hunter syndrome).";
 RL Hum. Mutat. 5:97-100(1995).
 [13]
 RN VARIANT MPS-II VAL-346.
 RP MEDLINE; 95322987.
 RX LI P., HUFFMAN P., THOMPSON J.N.;
 RA "Mutations of the iduronate-2-sulfatase gene on a T146T background in
 RT three patients with Hunter syndrome.";
 RL Hum. Mutat. 5:272-274(1995).
 [14]
 RN VARIANTS MPS-II.
 RP MEDLINE; 97094177.
 RX RATHMANN M., BUNGE S., BECK M., KRESSE H., TYLKI-SZYMANSKA A., GAL A.;
 RA "Mucopolysaccharidosis type II (Hunter syndrome): mutation 'hot spots'
 RT in the iduronate-2-sulfatase gene.";
 RL Am. J. Hum. Genet. 59:1202-1209(1996).

[15]
 RN VARIANTS MPS-II LEU-333 AND ASP-346.
 RX MEDLINE; 96163494.
 RA OLSEN T.C., EIKEN H.G., KNAPPSKOG P.M., KASE B.F., MANSSON J.-E.,
 RA BOMAN H., APOLD J.;
 RT "Mutations in the iduronate-2-sulfatase gene in five Norwegians with
 RL Hunter syndrome.";
 RL Hum. Genet. 97:198-203(1996).
 [16]
 RN VARIANTS MPS-II ASP-63; THR-347; GLN-468 AND LEU-468.
 RP MEDLINE; 97365936.
 RX VILLANI G.R.D., BALZANO N., GROSSO M., SALVADORE F., IZZO P.,
 RA DI NATALE P.;
 RA "Mucopolysaccharidosis type II: identification of six novel mutations
 RT in Italian patients.";
 RL Hum. Mutat. 10:71-75(1997).
 [17]
 RN VARIANT MPS-II GLN-468.
 RP MEDLINE; 98041699.
 RX SUKIGAWA K., SONG X.-Q., MASUNO M., FUKAO T., SHIMOZAWA N., FUKUDA S.,
 RA ISOGAI K., NISHIO H., MATSUO M., TOMATSU S., KONDO N., ORII T.;
 RA "Hunter disease in a girl caused by R468Q mutation in the
 RT iduronate-2-sulfatase gene and skewed inactivation of the X
 RL chromosome carrying the normal allele.";
 RL Hum. Mutat. 10:361-367(1997).
 [18]
 RN VARIANTS MPS-II N-45; Y-115; L-228; R-266; K-434; K-485 AND C-502.
 RP MEDLINE; 99092178.
 RX VAFIADAKI E., COOPER A., HEPTINSTALL L.E., HATTON C.E., THORNLEY M.,
 RA WRAITH J.E.;
 RT "Mutation analysis in 57 unrelated patients with MPS II.";
 RL Arch. Dis. Child. 79:237-241(1998).
 [19]
 RN VARIANTS MPS-II.
 RP MEDLINE; 99118886.
 RX KARSTEN S., VOSKOBOEVA E., TISHKANINA S., PETERSSON U.,
 RA KRASNOPOLSKAJA X., BONDESON M.-L.;
 RT "Mutational spectrum of the iduronate-2-sulfatase (IDS) gene in 36
 RL unrelated Russian MPS II patients.";
 RL Hum. Genet. 103:732-735(1998).
 [20]
 RN VARIANTS MPS-II LEU-86; HIS-88; PRO-88; ILE-118 AND HIS-266.
 RP BALZANO N., VILLANI G.R.D., GROSSO M., IZZO P., DI NATALE P.;
 RA "Detection of four novel mutations in the iduronate-2-sulfatase
 RT gene.";
 RL Hum. Mutat. 11:333-333(1998).
 [21]
 RN VARIANTS MPS-II THR-85; HIS-88; ILE-349 AND VAL-521.
 RP MEDLINE; 98112423.
 RX GORT L., COLL M.J., CHABAS A.;
 RT "Mutations in the iduronate-2-sulfatase gene in 12 Spanish patients
 RL with Hunter disease.";
 RL Hum. Mutat. Suppl. 1:S66-S68(1998).
 [22]
 RN VARIANTS MPS-II PHE-143; TRP-184; VAL-269 AND HIS-348.
 RP KARSTEN S.L., VOSKOBOEVA E., CARLBERG B.-M., KLEIJER W.J.,
 RA TOENNESSEN T., PETERSSON U., BONDESON M.-L.;
 RT "Identification of 9 novel gene mutations in 19 unrelated Hunter
 RL syndrome (Mucopolysaccharidosis type II) patients.";
 RL Hum. Mutat. 12:433-433(1998).
 CC -!- FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN
 CC SULFATE AND DERMATAN SULFATE.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE
 CC L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE
 CC AND HEPARIN.
 CC -!- SUBUNIT: LIVER IDS IS COMPOSED OF TWO MAJOR FORMS (A AND B) WHICH
 CC CONTAIN BOTH A 42 KD AND A 14 KD POLYPEPTIDES.
 CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.
 CC -!- TISSUE SPECIFICITY: LIVER, KIDNEY, LUNG, AND PLACENTA.
 CC -!- DISEASE: DEFECTS IN IDS ARE THE CAUSE OF HUNTER SYNDROME; AN X-
 CC LINKED INBORN ERROR ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE II
 CC (MPS-II); LEADING TO LYSOSOMAL ACCUMULATION OF HEPARAN SULFATE AND
 CC DERMATAN SULFATE AND THEIR EXCRETION IN URINE. MOST CHILDREN WITH

```
CC MPS-II HAVE A SEVERE FORM WITH EARLY SOMATIC ABNORMALITIES
CC INCLUDING SKELETAL DEFORMITIES, HEPATOSPLENOMEGALY, AND
CC PROGRESSIVE CARDIOPULMONARY DETERIORATION. A PROMINENT FEATURE IS
CC NEUROLOGICAL DAMAGE THAT PRESENTS AS DEVELOPMENTAL DELAY AND
CC HYPERACTIVITY BUT PROGRESSES TO MENTAL RETARDATION AND DEMENTIA.
CC THEY DIE BEFORE 15 YEARS OF AGE, USUALLY AS A RESULT OF
CC OBSTRUCTIVE AIRWAY DISEASE OR CARDIAC FAILURE. IN CONTRAST, THOSE
CC WITH A MILD FORM OF MPS-II MAY SURVIVE INTO ADULTHOOD, WITH
...: remainder of annotations omitted.

Query Match      100.0%; Score 4069; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPPRTGRLGLWGLVSSVCVAGLSETOANSTTDALNVLIIYDRLSPSLGCGDKLVR 60
QY 1 MPPRTGRLGLWGLVSSVCVAGLSETOANSTTDALNVLIIYDRLSPSLGCGDKLVR 60

Db 61 SPNIDQLASHSLLFONAFQAQVACAPSRVSLTGRPDTRLYDFNSVWRVHAGNFSTIP 120
QY 61 SPNIDQLASHSLLFONAFQAQVACAPSRVSLTGRPDTRLYDFNSVWRVHAGNFSTIP 120

Db 121 QYFKENGVTMSVGVKPHGPGISSNHTDSDPSYSPPPYHPSEKYEYNTKCRGPDGELHA 180
QY 121 QYFKENGVTMSVGVKPHGPGISSNHTDSDPSYSPPPYHPSEKYEYNTKCRGPDGELHA 180

Db 181 NLLCPVDVLDVPEGTLPDKQTEQAQLLEKMKTSASFFFLAVGYKPHIPPYPKFOK 240
QY 181 NLLCPVDVLDVPEGTLPDKQTEQAQLLEKMKTSASFFFLAVGYKPHIPPYPKFOK 240

Db 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNISVPYGPVDFQKIRQSY 300
QY 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNISVPYGPVDFQKIRQSY 300

Db 301 FASVSYLDQVGRLLSALDDLQANSTIIATSDHGALGEGHWAKYSNFDVATHVPLI 360
QY 301 FASVSYLDQVGRLLSALDDLQANSTIIATSDHGALGEGHWAKYSNFDVATHVPLI 360

Db 361 FYVPGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELSLFPTLAGLAGLQVPP 420
QY 361 FYVPGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELSLFPTLAGLAGLQVPP 420

Db 421 RCPVPSEHVELCRGKLLKHFRLDEEDPYLPGNPRELIAYSQYPRSDIPQWNSDKP 480
QY 421 RCPVPSEHVELCRGKLLKHFRLDEEDPYLPGNPRELIAYSQYPRSDIPQWNSDKP 480

Db 481 SLKDIKINGYSIRTDYRYTWVGFNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540
QY 481 SLKDIKINGYSIRTDYRYTWVGFNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540

Db 541 GGDLLQLIMP 550
QY 541 GGDLLQLIMP 550

RESULT 2
ID IDS_MOUSE STANDARD; PRT; 563 AA.
AC Q08890;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).
GN IDS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 93315172.
RA DANIELE A., FAUST C.J., HERMAN G.E., DI NATALE P., BALLABIO A.;
RT *Cloning and characterization of the cDNA for the murine iduronate
```

```
RT sulfatase gene.";
RL Genomics 16:755-757(1993).
CC FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN
CC Sulfate and DERMATAN SULFATE.
CC CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE
CC L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE
CC AND HEPARIN.
CC SUBCELLULAR LOCATION: LYSOSOMAL.
CC SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07921; AAA37880.1; -.
DR PIR; A47153; A47153.
DR MGD; MGI:96417; IDS.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PFAM; PF00884; Sulfatase; 1.
KW Hydrolase; Glycoprotein; Lysosome; zymogen; Signal.
FT SIGNAL 1 ? POTENTIAL.
FT PROPEP ? 59 BY SIMILARITY.
FT CHAIN 60 563 IDURONATE 2-SULFATASE.
FT MOD_RES 99 99 2-AMINO-3-OXOPROPIONIC ACID (BY
FT SIMILARITY).
FT CARBOHYD 130 130 POTENTIAL.
FT CARBOHYD 159 159 POTENTIAL.
FT CARBOHYD 261 261 POTENTIAL.
FT CARBOHYD 295 295 POTENTIAL.
FT CARBOHYD 526 526 POTENTIAL.
FT CARBOHYD 550 550 POTENTIAL.
SQ SEQUENCE 563 AA; 63437 MW; 9A6F7CD9 CRC32;

Query Match      76.9%; Score 3129; DB 1; Length 563;
Best Local Similarity 79.0%; Pred. No. 0.00e+00;
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps 4;

Db 19 ARAIWQLSFLSLGSECTALESAAAGNSATDALNILLIIVDDLRLPSLGCYDKLVRSPN 78
QY 4 PRTGRLLWGLVSSVCVAGLSETOANSTTDALNVLIIYDRLSPSLGCGDKLVRSPN 63

Db 79 IDOLASHSLVLFONAFQAQVACAPSRVSLTGRPDTRLYDFNSVWRVHAGNFSTIPQYF 138
QY 64 IDOLASHSLVLFONAFQAQVACAPSRVSLTGRPDTRLYDFNSVWRVHAGNFSTIPQYF 123

Db 139 KENGYVTMSVGVKPHGPGISSNHTDSDPSYSPPPYHPSEKYEYNTKCRGDKLHANLL 198
QY 124 KENGYVTMSVGVKPHGPGISSNHTDSDPSYSPPPYHPSEKYEYNTKCRGDKLHANLL 183

Db 199 CPVDVADVPEGTLPDKQSTEEAIRLLEKMKTSPPFLAVGYKPHIPPYPKFOKLYP 258
QY 184 CPVDVADVPEGTLPDKQSTEEAIRLLEKMKTSPPFLAVGYKPHIPPYPKFOKLYP 243

Db 259 LENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNISVPYGPVDFQKIRQSYFAS 318
QY 244 LENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNISVPYGPVDFQKIRQSYFAS 303

Db 319 VSYLDQVGHVLSALDDLQANSTIIATSDHGALGEGHWAKYSNFDVATHVPLMLYV 378
QY 304 VSYLDQVGHVLSALDDLQANSTIIATSDHGALGEGHWAKYSNFDVATHVPLMLYV 363

Db 379 PGRTAPLPAAGKLFYPRDPFDPASDMMADGRTEDLVELVSLFPTLAGLAGLVLGAP 438
QY 364 PGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELVSLFPTLAGLAGLQVPRCP 423

Db 439 SLLFMLSFAEKAR-IFRSTC-SSMTWKRSTCLVPRELIAYSQYPRPADPQWNSDKPTL 496
QY 424 VPSFHVLCREGKLLKHFRLDE-EDPYLPGNPRELIAYSQYPRSDIPQWNSDKPSL 482
```

```
Db 497 NDIRSWDIYH-VDYRYTVVWGFDPSEFLANFSDIHAGELYFVSDPDLQDNHNVYNDSSHG 555
   :||: :|||||:| :|||||:| :|||||:| :|||||:| :|||||:| :|||||:| :|||||:|
QY 483 KDIKIMGYSIRTIDYRYTVVWGFDPSEFLANFSDIHAGELYFVSDPDLQDNHNVYNDSSGG 542
   :||: :|||||:| :|||||:| :|||||:| :|||||:| :|||||:| :|||||:| :|||||:|

Db 556 GLLHSLRP 563
   :||: :||:|
QY 543 DLFOLLMP 550
   :||: :||:|

RESULT 3
ID YIDJ_ECOLI STANDARD; PRT; 497 AA.
AC F31447;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 57.3 KD PROTEIN IN EMRD-GLVG INTERGENIC REGION.
GN YIDJ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RL genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
CC -|- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L10328; AAA62030.1; -.
DR EMBL; AE000445; AAC76701.1; -.
DR ECOGENE; EG11705; YIDJ.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PFAM; PF00884; Sulfatase; 1.
KW Hypothetical protein; Hydrolase.
FT ACT_SITE 102 102 POTENTIAL.
SQ SEQUENCE 497 AA; 57295 MW; 9D12470F CRC32;

Query Match 6.9%; Score 282; DB 1; Length 497;
Best Local Similarity 28.5%; Pred. No. 6.79e-34;
Matches 94; Conservative 83; Mismatches 125; Indels 28; Gaps 21;

Db 5 NLFVMTDTQATNMVGYSGKPLNTQNTIDSLAEGIRFNSAYTCSPVCTPARAGLTGIY 64
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 38 NVLLIIVDLRPSL-GCYGDKLVSPNIDQLASHLLFQNAFAQAQVCAVPSRVSLTCRR 96
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

Db 65 ANOSGPTW-NV--APGKNTSTMGRYFKDAGYHTCYGK-WHLGDHDFYFGEGCEPPWD- 119
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 97 PDTRLDYFNSYWRVHAGNFSTIPQYFKNGYVMTSGVKVPH-PGISSNHTDDSPYSWSF 155
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

Db 120 ADYWFDCANVLSLT--EKEISLWRNGLSNVEDLQANHIDETFTTWAHRISNRAVDFLQOP 177
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 156 PPVHPSEKVENTKTCRGPDELHANLCPVDVLDV-P-EGTLP-DKOSTQOALQLEKM 212
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

Db 178 ARADEPEPLMVSYDEPHPTCTPVEYLEKYA-D-FYVELGEKAODL---ANKP--E-HH 229
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 213 KTSAPPEFLAVGKHPHPFYKPEFKLYPLENITLAPDEVPDGLPPVAINPMDIRQ 272
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

Db 230 RLWAQAMP-S-PVGDDGL-YHHL----YFACNDFVDQIGRVINALTPEQRENTWVI-YT 282
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 273 REDVQALNISVPYGPVPVDFORKIRQSYFASVSYLDTQVGRLLSALDQLANSTIIAFT 332
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
```

```
Db 283 SDHGEMGAHKLSKGAAMYDDITRIPLII 312
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 333 SDHGALGEHGEWAKYSN-FDVATHVPLIF 361
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 4
ID STS_MOUSE STANDARD; PRT; 624 AA.
AC P50427;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE STERYL-SULFATASE PRECURSOR (EC 3.1.6.2) (STEROID SULFATASE) (STERYL-
DE SULFATE SULFOHYDROLASE) (ARYLSULFATASE C) (ASC).
GN STS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE; 96241577.
RA SALIDO E.C., LI X.M., YEN P.H., MARTIN N., MOHANDAS T.K.,
RA SHAPIRO L.J.;
RT "Cloning and expression of the mouse pseudoautosomal steroid
RT sulphatase gene (Sts).";
RL Nat. Genet. 13:83-86(1996).
CC -|- FUNCTION: CONVERSION OF SULFATED STEROID PRECURSORS TO ESTROGENS
CC DURING PREGNANCY.
CC -|- CATALYTIC ACTIVITY: 3-BETA-HYDROXYANDROST-5-EN-17-ONE 3-SULFATE +
CC H(2)O = 3-BETA-HYDROXYANDROST-5-EN-17-ONE + SULFATE.
CC -|- SUBUNIT: HOMODIMER.
CC -|- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE. THE SEQUENCE SHOWS
CC SEVERAL MEMBRANE-SPANNING DOMAINS THAT COULD SERVE TO ANCHOR THE
CC PROTEIN IN THE MICROSOMAL MEMBRANE.
CC -|- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U37545; AAB09308.1; -.
DR HSSP; P15848; IFSU.
DR MGD; MGI:98438; STS.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PFAM; PF00884; Sulfatase; 1.
KW Hydrolase; Microsome; Transmembrane; Glycoprotein; Steroid metabolism;
KW Pregnancy; Signal.
FT SIGNAL 1 ? POTENTIAL.
FT CHAIN 1 ? 624 2-AMINO-3-OXOPROPIONIC ACID (BY
FT MOD_RES 83 83 SIMILARITY).
FT ACT_SITE 144 144 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT CARBOHYD 55 55 POTENTIAL.
FT CARBOHYD 465 465 POTENTIAL.
SQ SEQUENCE 624 AA; 66590 MW; 0A371AC7 CRC32;

Query Match 6.5%; Score 266; DB 1; Length 624;
Best Local Similarity 26.8%; Pred. No. 1.50e-30;
Matches 118; Conservative 116; Mismatches 164; Indels 43; Gaps 37;

Db 2 PRP-RPLL-LAVMAATLADIILAAADPAPAPAPRPPNFFLLIMADDIGDLCYGNKTLR 59
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 4 PRTGGLLWGLVLSVVC-VALGSE-TQANSTTDALNVLIIIVDDL-RPSLGCYGDKLVR 60
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

Db 60 TPHLDRLAREGVKLTQHLAAAPLCTPSRAFLTGRYPFRSGMAAHGVRGVYLFTASSGGL 119
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
```

QY 61 SPNIDQLASHLLFQNAFAQAQVACAPSRVSFELTGR-RDPTTRL-YD-FNSY-WRVHAGNF 116

Db 120 PPSEVTMARLLKGRYATALICK-WHLGLSCRGATDFCHHPLRHGDFRPLGVPTT--NLR 176

QY 117 --S--TIQYKENGCVTMSGVKVPHPGISS-NHTD--DSPYSWSPFPVHPSPSEKVENTK 169

Db 177 DCRGAGTVFGALRVFAAGPLAALGASLAMAARWAGLARVPGWALAGTAAMLAAGG 236

QY 170 TCR-GPDGELHANL-LCPVDVLDVPEGLTPDKQSTE-QAI-QLLEKMKTSASFPFLAVGY 225

Db 237 PRSASCLGF-RP-ANCFMLDDLAQRDTYGLLFRRLADEAALFLRRNRPFLFLSLF 294

QY 226 HKPHIPFPYKREFQKLYPLENITLAPDEVPDGLPP-VAYNPWMDIRQ-RE-D-VQALN- 280

Db 295 LHVHTAHADPGFAGRSUHGAYGDSVEMDGCVGLAALDELGLARETLVYFTSDHGAH 354

QY 281 ISVPYGPPI-VDFQ-RKJQSTFASVSYLDVQVGRLLSALDQLANSTIIAFTSDHGA 338

Db 355 VEELGPRGE--RMG-GSN-G-VF-RGGKGNW-EGGVRY-PCLVRW--PREL-SPGRVVA 403

QY 339 LGHEGAWAKSNFVATHVPLIFYVPGRTASLPEAGEKLFVYLDPDFDSASQLMEGRQSM 398

Db 404 EPTSLMDVFPVARLAGAELP 424

QY 399 DLVELVSLFPTLAGLAGLQVP 419

RESULT 5

ID ARSF_HUMAN STANDARD; PRT; 591 AA.

AC P54793;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ARYLSULFATASE F PRECURSOR (EC 3.1.6.-) (ASF).

GN ARSF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-KIDNEY;

RX MEDLINE; 95236447.

RA FRANCO B., MERONI G., PARENTI G., LEVILLIERS J., BERNARD L.,

RA GEBBIA M., COX L., MAROTEAUX P., SHEFFIELD L., RAPPOLD G.A.,

RA ANDRIA G., PETIT C., BALLABIO A.;

RT "A cluster of sulfatase genes on xp22.3: mutations in

RT chondrodysplasia punctata (CDPX) and implications for warfarin

RT embryopathy.";

RL Cell 81:15-25(1995).

CC -1- ENZYME REGULATION: NOT INHIBITED BY DHEAS OR WARFARIN.

CC -1- MISCELLANEOUS: OPTIMUM PH IS 8.

CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X97868; CAA66462.1; -

DR HSSP; P15289; IAUK.

MM; 300003; -

DR PROSITE; PS00523; SULFATASE_1; 1.

DR PROSITE; PS00149; SULFATASE_2; 1.

KW Hydrolase; Glycoprotein; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 591 ARYLSULFATASE F.

FT MOD_RES 79 79 2-AMINO-3-OXOPROPIONIC ACID (BY

FT ACT_SITE 140 140 BY SIMILARITY.

FT CARBOHYD 51 51 POTENTIAL.

FT CARBOHYD 118 118 POTENTIAL.

FT CARBOHYD 337 337 POTENTIAL.

SQ SEQUENCE 591 AA; 66003 MM; 233649DS CRC32;

Query Match 6.38; Score 258; DB 1; Length 591;

Best Local Similarity 27.98; Pred. No. 6.75e-29;

Matches 97; Conservative 86; Mismatches 133; Indels 32; Gaps 24;

Db 7 LVFMSLVCALLNTPWGHTGCMTRPNIVLIWDDLGIGDLCYGNDRTPHIDRLAREG 66

QY 15 LV-LSSVCVALGS-ETQANSTTDALNVLIIIVDDL-RPSCCYGDKLVRSFNIDQLASHS 71

Db 67 VRLTQHISRAASLCSPRSFAFLTGRYPISQWVS-SGNRRVIONLAVPAGLPLNETTLAAL 125

QY 72 LLFQNAFAQAQVACAPSRVSFELTGRPDTRLYDFNSYRV-H-----AG-NF--STIPQY 122

Db 126 LKQGYSTGLIGK-WHQGLNCDRSDOCHHPYNGFYDYGMPTFLVDSQWPDSPRNTL 184

QY 123 FKENGVTMSVGVFHPGISSN-HTDSD--PYSWSFPPHPSPSEKY-EN--TKTCRGPDG 176

Db 185 AFESQWLCL-VOLVAITAILTFTKLSGWVSVPLLIIFSLIFLAGYAWFSHTSPLY 243

QY 177 ELHANL-LCPVDVLDVPEGLTP-DKOSTEQAIQLLEKMKTSASFPFLAVGHKPHIPFY 234

Db 244 -WDCL-LMRGHEITEQPMKAERAGSIMVKEAISPLE-RHSKETFLFLFSLVHTPLPTT 300

QY 235 PKFEQKLYPLENITLAP-DEVPDGLPPVAYNPWMDIRQEDVQALNIS-VP-YGPIPV- 290

Db 301 DDFTGTSKHGLGDNVEEMDSWVGKILDAIDDFGLRNNTLYVFTSDHG 348

QY 291 -DFQRKIQRS-YEASVSYLDVQVGRLLSALDQLANSTIIAFTSDHG 336

RESULT 6

ID ARSF_HUMAN STANDARD; PRT; 589 AA.

AC P51690;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ARYLSULFATASE E PRECURSOR (EC 3.1.6.-) (ASE).

GN ARSF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A., AND VARIANTS CDPX.

RC TISSUE-KIDNEY;

RX MEDLINE; 95236447.

RA FRANCO B., MERONI G., PARENTI G., LEVILLIERS J., BERNARD L.,

RA GEBBIA M., COX L., MAROTEAUX P., SHEFFIELD L., RAPPOLD G.A.,

RA ANDRIA G., PETIT C., BALLABIO A.;

RT "A cluster of sulfatase genes on xp22.3: mutations in

RT chondrodysplasia punctata (CDPX) and implications for warfarin

RT embryopathy.";

RL Cell 81:15-25(1995).

RN [2]

RP VARIANT CDPX TVR-492.

RX MEDLINE; 98072521.

RA PARENTI G., BUTTITA P., MERONI G., FRANCO B., BERNARD L.,

RA RIZZOLO M.G., BRUNETTI-PIERRI N., BALLABIO A., ANDRIA G.;

RT "X-linked recessive chondrodysplasia punctata due to a new point

RT mutation of the ARSE gene.";

RL Am. J. Med. Genet. 73:139-143(1997).

AC P34059;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE N-ACETYLALACTOSAMINE-6-SULFATASE PRECURSOR (EC 3.1.6.4) (N-
DE ACETYLALACTOSAMINE-6-SULFATE SULFATASE) (GALACTOSE-6-SULFATE
DE SULFATASE) (GALNAC6S SULFATASE) (CHONDROITINSULFATASE)
DE (CHONDROITINASE).
GN GALNS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=PLACENTA;
RX MEDLINE; 92095973.
RA TOMATSU S., FUKUDA S., MASUE M., SUKEGAWA K., FUKAO T., YAMAGISHI A.,
RA HORI T., IWATA H., OGAWA T., NAKASHIMA Y., HANYU Y., HASHIMOTO T.,
RA TITANI K., OYAMA R., SUZUKI M., YAGI K., HAYASHI Y., ORII T.;
RT "Morquio disease: isolation, characterization and expression of full-
RT length cDNA for human N-acetylglactosamine-6-sulfate sulfatase.";
RT Biochem. Biophys. Res. Commun. 181:677-683(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 95095267.
RA MORRIS C.P., GUO X.H., APOSTOLOU S., HOPWOOD J.J., SCOTT H.S.;
RT "Morquio A syndrome: cloning, sequence, and structure of the human N-
RT acetylglactosamine 6-sulfatase (GALNS) gene.";
RL Genomics 22:652-654(1994).
RN [3]
RN VARIANT MPS-IVA LYS-204.
RX MEDLINE; 92395122.
RA FUKUDA S., TOMATSU S., MASUE M., SUKEGAWA K., IWATA H., OGAWA T.,
RA NAKASHIMA Y., HORI T., YAMAGISHI A., HANYU Y., MOROOKA K., KIMAN T.,
RA HASHIMOTO T., ORII T.;
RT "Mucopolysaccharidosis type IVA. N-acetylglactosamine-6-sulfate
RT sulfatase exonic point mutations in classical Morquio and mild
RT cases.";
RT J. Clin. Invest. 90:1049-1053(1992).
RN [4]
RN VARIANTS MPS-IVA.
RX MEDLINE; 95397840.
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., MARUF REZVI G.,
RA YAMAGISHI A., YAMADA N., KATO Z., ISOGAI K., SUKEGAWA K., KONDO N.,
RA SUZUKI Y., SHIMOZAWA N., ORII T.;
RT "Mucopolysaccharidosis IVA: identification of a common missense
RT mutation I113F in the N-Acetylglactosamine-6-sulfate sulfatase
RT gene.";
RL Am. J. Hum. Genet. 57:556-563(1995).
RN [5]
RN VARIANTS MPS-IVA.
RX MEDLINE; 95315929.
RA OGAWA T., TOMATSU S., FUKUDA S., YAMAGISHI A., MARUF REZVI G.,
RA SUKEGAWA K., KONDO N., SUZUKI Y., SHIMOZAWA N., ORII T.;
RT "Mucopolysaccharidosis IVA: screening and identification of mutations
RT of the N-acetylglactosamine-6-sulfate sulfatase gene.";
RL Hum. Mol. Genet. 4:341-349(1995).
RN [6]
RN VARIANTS MPS-IVA VAL-77; TRP-90; VAL-96; LEU-151; GLY-230 & THR-291.
RX MEDLINE; 95359983.
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., MARUF REZVI G.,
RA YAMAGISHI A., YAMADA N., KATO Z., ISOGAI K., SUKEGAWA K., KONDO N.,
RA SUZUKI Y., SHIMOZAWA N., ORII T.;
RT "Mucopolysaccharidosis type IVA: identification of six novel
RT mutations among non-Japanese patients.";
RL Hum. Mol. Genet. 4:741-743(1995).
RN [7]
RN VARIANT MPS-IVA SER-487.
RX MEDLINE; 96047158.
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., YAMADA N., ISOGAI K.,
RA KATO Z., SUKEGAWA K., KONDO N., SUZUKI Y., SHIMOZAWA N., ORII T.;
RT "Two new mutations, Q473X and N487S, in a Caucasian patient with
RT mucopolysaccharidosis IVA (Morquio disease).";

Hum. Mutat. 6:195-196(1995).
[8]
RN VARIANTS MPS-IVA ALA-138; SER-151 AND LEU-151.
RX MEDLINE; 96216146.
RA TOMATSU S., FUKUDA S., YAMAGISHI A., COOPER A., WRAITH J.E., HORI T.,
RA KATO Z., YAMADA N., ISOGAI K., SUKEGAWA K., KONDO N., SUZUKI Y.,
RA SHIMOZAWA N., ORII T.;
RT "Mucopolysaccharidosis IVA: four new exonic mutations in patients
RT with N-acetylglactosamine-6-sulfate sulfatase deficiency.";
RL Am. J. Hum. Genet. 58:950-962(1996).
RN [9]
RN VARIANTS MPS-IVA CYS-94 AND VAL-97.
RX MEDLINE; 96423834.
RA COLE D.E.C., FUKUDA S., GORDON B.A., RIP J.W., LECOUTEUR A.N.,
RA RUPAR C.A., TOMATSU S., OGAWA T., SUKEGAWA K., ORII T.;
RT "Heteroallelic missense mutations of the galactosamine-6-sulfate
RT sulfatase (GALNS) gene in a mild form of Morquio disease (MPS IVA).";
RL Am. J. Med. Genet. 63:558-565(1996).
RN [10]
RN VARIANTS MPS-IVA.
RX MEDLINE; 97442274.
RA BUNGE S., KLEIJER W.J., TYLKI-SZYMANSKA A., STEGLICH C., BECK M.,
RA TOMATSU S., FUKUDA S., POORTHUIS B.J.H.M., CZARTORYSKA B., ORII T.,
RA GAL A.;
RT "Identification of 31 novel mutations in the N-acetylglactosamine-6-
RT sulfatase gene reveals excessive allelic heterogeneity among patients
RT with Morquio A syndrome.";
RL Hum. Mutat. 10:223-232(1997).
RN [11]
RN VARIANTS MPS-IVA.
RX MEDLINE; 98041700.
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., FERREIRA P.,
RA DI NATALE P., TORTORA P., FUJIMOTO A., KATO Z., YAMADA N., ISOGAI K.,
RA YAMAGISHI A., SUKEGAWA K., SUZUKI Y., SHIMOZAWA N., KONDO N.,
RA SLY W.S., ORII T.;
RT "Fourteen novel mucopolysaccharidosis IVA producing mutations in
RT GALNS gene.";
RL Hum. Mutat. 10:368-375(1997).
RN [12]
RN VARIANTS MPS-IVA.
RX MEDLINE; 98180718.
RA YAMADA N., FUKUDA S., TOMATSU S., MULLER V., HOPWOOD J.J., NELSON J.,
RA KATO Z., YAMAGISHI A., SUKEGAWA K., KONDO N., ORII T.;
RT "Molecular heterogeneity in mucopolysaccharidosis IVA in Australia
RT and Northern Ireland: nine novel mutations including T312S, a common
RT allele that confers a mild phenotype.";
RL Hum. Mutat. 11:202-208(1998).
RN [13]
RN VARIANTS SER-393 AND MET-488.
RX MEDLINE; 98112415.
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., YAMAGISHI A., KATO Z.,
RA YAMADA N., ISOGAI K., SUKEGAWA K., SUZUKI Y., SHIMOZAWA N., KONDO N.,
RA ORII T.;
RT "Fifteen polymorphisms in the N-acetylglactosamine-6-sulfate
RT sulfatase (GALNS) gene: diagnostic implications in Morquio disease.";
RL Hum. Mutat. Suppl. 1:S42-S46(1998).
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 6-SULFATE GROUPS OF THE
CC N-ACETYL-D-GALACTOSAMINE 6-SULFATE UNITS OF CHONDROITIN SULFATE
CC AND OF THE D-GALACTOSE 6-SULFATE UNITS OF KERATAN SULFATE.
CC -!- SUBUNIT: OLIGOMER OF DISULFIDE LINKED 40- AND 15-KD POLYPEPTIDES.
CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC -!- DISEASE: DEFECTS IN GALNS ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS
CC TYPE IVA (MPS-IVA) (ALSO KNOWN AS MORQUIO A SYNDROME) WHICH IS
CC CHARACTERIZED BY SPECIFIC SPONDYLOEPHYSEAL DYSPLASIA, SHORT
CC TRUNK DWARFISM, COXA VALGA, ODONTOID HYPOPLASIA, CORNEAL
CC OPACITIES, PRESERVATION OF INTELLIGENCE, AND EXCESSIVE URINARY
CC EXCRETION OF KERATAN SULFATE AND CHONDROITIN-6-SULFATE. SEVERELY
CC AFFECTED PATIENTS USUALLY DIE OF CARDIOPULMONARY DISTURBANCE OR
CC CERVICAL CORD COMPRESSION IN THE SECOND OR THIRD DECADE OF LIFE.
CC -!- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARYLISULFATASE A PRECURSOR (EC 3.1.6.8) (ASA) (CEREBROSIDE-SULFATASE).
GN ARSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90361046.
RA KREYSING J., VON FIGURA K., GIESELMANN V.;
RT "Structure of the arylsulfatase A gene.";
RL Eur. J. Biochem. 191:627-631(1990).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 89093115.
RA STEIN C., GIESELMANN V., KREYSING J., SCHMIDT B., POHLMANN R.,
WAHEED A., MEYER H.E., O'BRIEN J.S., VON FIGURA K.;
RT "Cloning and expression of human arylsulfatase A";
RL J. Biol. Chem. 264:1252-1259(1989).
RN [3]
RN SEQUENCE FROM N.A.
RA ADAMS M.D., KERLAVAGE A.R., FULDNER R.A., PHILLIPS C.A., VENTER J.C.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 19-33 AND 434-479, AND SUBUNITS.
RX MEDLINE; 92338230.
RA FUJII T., KOBAYASHI T., HONKE K., GASA S., ISHIKAWA M., SHIMITSU T.,
RA MAKITA A.;
RT "Proteolytic processing of human lysosomal arylsulfatase A";
RL Biochim. Biophys. Acta 1122:93-98(1992).
RN [5]
RN PARTIAL SEQUENCE, AND 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
RX MEDLINE; 95354208.
RA SCHMIDT B., SELMER T., INGENDOH A., VON FIGURA K.;
RT "A novel amino acid modification in sulfatases that is defective in
multiple sulfatase deficiency.";
RL Cell 82:271-278(1995).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE; 98191325.
RA LUKATELA G., KRAUSS N., THEIS K., SELMER T., GIESELMANN V.,
RA VON FIGURA K., SAENGER W.;
RT "Crystal structure of human arylsulfatase A: the aldehyde function
and the metal ion at the active site suggest a novel mechanism for
sulfate ester hydrolysis.";
RL Biochemistry 37:3654-3664(1998).
RN [7]
RN REVIEW ON MLD VARIANTS.
RX MEDLINE; 95170731.
RA GIESELMANN V., ZLOTOGORA J., HARRIS A., WENGER D.A., MORRIS C.P.;
RT "Molecular genetics of metachromatic leukodystrophy.";
RL Hum. Mutat. 4:233-242(1994).
RN [8]
RN VARIANT MLD GLN-84.
RX MEDLINE; 92344341.
RA KAPPLER J., VON FIGURA K., GIESELMANN V.;
RT "Late-onset metachromatic leukodystrophy: molecular pathology in two
siblings.";
RL Ann. Neurol. 31:256-261(1992).
RN [9]
RN VARIANT MLD PHE-96.
RX MEDLINE; 91328147.
RA GIESELMANN V., FLUHARTY A.L., TONNESSEN T., VON FIGURA K.;
RT "Mutations in the arylsulfatase A pseudodeficiency allele causing
metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 49:407-413(1991).
RN [10]
RN VARIANT MLD ASP-99.
RX MEDLINE; 91206410.
RA KONDO R., WAKAMATSU N., YOSHINO H., FUKUHARA N., MIYATAKE T.,
RA TSUJII S.;
RT "Identification of a mutation in the arylsulfatase A gene of a
patient with adult-type metachromatic leukodystrophy.";
RL

RL Am. J. Hum. Genet. 48:971-978(1991).
RN [11]
RN VARIANT MLD LEU-426, AND VARIANTS CYS-193 AND SER-391.
RX MEDLINE; 91074201.
RA POLTEN A., FLUHARTY A.L., FLUHARTY C.B., KAPPLER J., VON FIGURA K.,
RA GIESELMANN V.;
RT "Molecular basis of different forms of metachromatic leukodystrophy.";
RL New Engl. J. Med. 324:18-22(1991).
RN [12]
RN VARIANT MLD SER-122.
RX MEDLINE; 94063853.
RA HONKE K., KOBAYASHI T., FUJII T., GASA S., XU M., TAKAMARU Y.,
RA KONDO R., TSUJII S., MAKITA A.;
RT "An adult-type metachromatic leukodystrophy caused by substitution of
serine for glycine-122 in arylsulfatase A.";
RL Hum. Genet. 92:451-456(1993).
RN [13]
RN VARIANT MLD LEU-136.
RX MEDLINE; 95163939.
RA KAFERT S., HEINISCH U., ZLOTOGORA J., GIESELMANN V.;
RT "A missense mutation P136L in the arylsulfatase A gene causes
instability and loss of activity of the mutant enzyme.";
RL Hum. Genet. 95:201-204(1995).
RN [14]
RN VARIANT MLD ARG-245.
RX MEDLINE; 93319632.
RA HASEGAWA Y., KAWANE H., ETO Y.;
RT "Mutations in the arylsulfatase A gene of Japanese patients with
metachromatic leukodystrophy.";
RL DNA Cell Biol. 12:493-498(1993).
RN [15]
RN VARIANT MLD MET-274.
RX MEDLINE; 94004907.
RA HARVEY J.S., NELSON P.V., CAREY W.F., ROBERTSON E.F., MORRIS C.P.;
RT "An arylsulfatase A (ARSA) missense mutation (T274M) causing late-
infantile metachromatic leukodystrophy.";
RL Hum. Mutat. 2:261-267(1993).
RN [16]
RN VARIANT MLD SER-309.
RX MEDLINE; 93318834.
RA KREYSING J., BOHNE W., BOSENBERG C., MARCHESINI S., TURPIN J.C.,
RA BAUMANN N., VON FIGURA K., GIESELMANN V.;
RT "High residual arylsulfatase A (ARSA) activity in a patient with
late-infantile metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 53:339-346(1993).
RN [17]
RN VARIANT SER-350.
RX MEDLINE; 90083282.
RA GIESELMANN V., POLTEN A., KREYSING J., VON FIGURA K.;
RT "Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal
and N-glycosylation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9436-9440(1989).
RN [18]
RN VARIANT MLD LEU-426.
RX MEDLINE; 93202658.
RA BARTH M.L., FENSOM A., HARRIS A.;
RT "Prevalence of common mutations in the arylsulfatase A gene in
metachromatic leukodystrophy patients diagnosed in Britain.";
RL Hum. Genet. 91:73-77(1993).
RN [19]
RN VARIANTS MLD VAL-212; VAL-224 AND TYR-295.
RX MEDLINE; 94154687.
RA BARTH M.L., FENSOM A., HARRIS A.;
RT "Missense mutations in the arylsulfatase A genes of metachromatic
leukodystrophy patients.";
RL Hum. Mol. Genet. 2:2117-2121(1993).
RN [20]
RN VARIANTS MLD LEU-82; TYR-172; CYS-201; GLN-311; VAL-335 AND TRP-390.
RX MEDLINE; 96047150.
RA BARTH M.L., FENSOM A., HARRIS A.;
RT "Identification of seven novel mutations associated with
metachromatic leukodystrophy.";
RL Hum. Mutat. 6:170-176(1995).

RN [21] CHARACTERIZATION OF VARIANTS MET-274 AND VAL-335.
 RP MEDLINE; 96303701.
 RA HESS B., KAPERT S., HEINISCH U., WENGER D.A., ZLOTOGORA J.,
 RA GIESELMANN V.;
 RT "Characterization of two arylsulfatase A missense mutations D335V and
 RT T274M causing late infantile metachromatic leukodystrophy.";
 RL Hum. Mutat. 7:311-317(1996).
 RN [22]
 RP VARIANTS MLD.
 RP MEDLINE; 97245886.
 RA DRAGHTA R., LETOURNEUR F., DRUGAN C., MANICOM J., BLANCHOT C.,
 RA KAHN A., POENARU L., CAILLAUD C.;
 RT "Metachromatic leukodystrophy: identification of the first deletion in
 RT exon 1 and of nine novel point mutations in the arylsulfatase A
 RT gene.";
 RL Hum. Mutat. 9:234-242(1997).
 RN [23]
 RP VARIANT MLD SER-406--THR-408 DEL.
 RX MEDLINE; 98141126.
 RA REGIS S., FILOCAMO M., STROPPIANO M., CORSOLINI F., CAROLI F.,
 RA GATTI R.;
 RT "A 9-bp deletion (2320del9) on the background of the arylsulfatase A
 RT pseudodeficiency allele in a metachromatic leukodystrophy patient and
 RT in a patient with nonprogressive neurological symptoms.";
 RL Hum. Genet. 102:50-53(1998).
 RN [24]
 RP VARIANTS MLD LEU-135 AND SER-179.
 RP MEDLINE; 98260871.
 RA GOMEZ-LIRA M., PERUSI C., MOTTES M., PIGNATTI P.F., MANFREDI M.,
 RA RIZZUTO N., SALVIATI A.;
 RT "Molecular genetic characterization of two metachromatic
 RT leukodystrophy patients who carry the T799G mutation and show
 RT different phenotypes: description of a novel null-type mutation.";
 RL Hum. Genet. 102:459-463(1998).
 RN [25]
 RP VARIANTS MLD GLN-390 AND TYR-397.
 RP MEDLINE; 98112481.
 RA COULTER-MACKIE M.B., GAGNIER L.;
 RT "Two novel mutations in the arylsulfatase A gene associated with
 RT juvenile (R390Q) and adult onset (H397I) metachromatic
 RT leukodystrophy.";
 RL Hum. Mutat. Suppl. 1:S254-S256(1998).
 RN [26]
 RP VARIANT HIS-496.
 RP MEDLINE; 98415722.
 RA RICKETTS M.H., PORTEZ R.D., MANOWITZ P.;
 RT "The R496H mutation of arylsulfatase A does not cause metachromatic
 RT leukodystrophy.";
 RL Hum. Mutat. 12:238-239(1998).

...
 Note: remainder of annotations omitted.

Query Match 4.8%; Score 195; DB 1; Length 507;
 Best Local Similarity 43.3%; Pred. No. 2.30e-16;
 Matches 29; Conservative 15; Mismatches 22; Indels 1; Gaps 1;

Db 22 NVILFADLLGCGDLCYGHPSSTPNLDQLAAGLRFTHFVPSVSLCTPSRAALLTGRLL 81
 :::: ||| |||| :|||::: ||: :|||: |||: :|||: |||:

QY 38 NVLLIIVDDLRL-PSLGCYGDKLVRSPNIDQLASHLLFNAFAQAQVCAPSRVSLTGR 96
 :||: ||| |||| :|||::: ||: :|||: |||: :|||: |||:

Db 82 PVRMGY 88
 | :|

QY 97 PDTRLY 103

RESULT 14

ID ARSB_HUMAN STANDARD; PRT; 533 AA.
 AC P15848;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE ARYLSULFATASE B PRECURSOR (EC 3.1.6.12) (ASB) (N-ACETYL GALACTOSAMINE-

DE 4-SULFATASE) (G4S).
 GN ARSB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 90153994.
 RA VINGRON M., MEYER H.E., POHLMANN R., VON FIGURA K.;
 RT "Phylogenetic conservation of arylsulfatases. cDNA cloning and
 RT expression of human arylsulfatase B.";
 RL J. Biol. Chem. 265:3374-3381(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE; 90152677.
 RA SCHUCHMAN E.H., JACKSON C.E., DESNICK R.J.;
 RT "Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-
 RT length cDNA, and regions of amino acid identity with arylsulfatases A
 RT and C.";
 RL Genomics 6:149-158(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93326648.
 RA MODARESSI S., RUPP K., VON FIGURA K., PETERS C.;
 RT "Structure of the human arylsulfatase B gene.";
 RL Biol. Chem. Hoppe-Seyler 374:327-335(1993).
 RN [4]
 RP SEQUENCE OF 1-104 FROM N.A.
 RX MEDLINE; 92028992.
 RA LITJENS T., MORRIS C.P., GIBSON G.J., BECKMANN K.R., HOPWOOD J.J.;
 RT "Human N-acetylgalactosamine-4-sulphatase: protein maturation and
 RT isolation of genomic clones.";
 RL Biochem. Int. 24:209-215(1991).
 RN [5]
 RP 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
 RP MEDLINE; 95354208.
 RA SCHMIDT B., SELMER T., INGENDOH A., VON FIGURA K.;
 RT "A novel amino acid modification in sulfatases that is defective in
 RT multiple sulfatase deficiency.";
 RL Cell 82:271-278(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE; 97184692.
 RA BOND C.S., CLEMENTS P.R., ASHEY S.J., COLLIER C.A., HARROP S.J.,
 RA HOPWOOD J.J., GUSS J.M.;
 RT "Structure of a human lysosomal sulfatase.";
 RL Structure 5:277-289(1997).
 RN [7]
 RP VARIANTS MPS-VI ARG-117; PRO-236 AND TYR-405.
 RX MEDLINE; 92197625.
 RA JIN W.-D., JACKSON C.E., DESNICK R.J., SCHUCHMAN E.H.;
 RT "Mucopolysaccharidosis type VI: identification of three mutations in
 RT the arylsulfatase B gene of patients with the severe and mild
 RT phenotypes provides molecular evidence for genetic heterogeneity.";
 RL Am. J. Hum. Genet. 50:795-800(1992).
 RN [8]
 RP VARIANT MPS-VI GLY-137, AND VARIANT MET-376.
 RX MEDLINE; 92042029.
 RA WICKER G., PRILL V., BROOKS D., GIBSON G., HOPWOOD J.,
 RA VON FIGURA K., PETERS C.;
 RT "Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome). An intermediate
 RT clinical phenotype caused by substitution of valine for glycine at
 RT position 137 of arylsulfatase B.";
 RL J. Biol. Chem. 266:21386-21391(1991).
 RN [9]
 RP VARIANTS MPS-VI MET-92; GLN-95; CYS-210; PRO-393 AND PRO-498.
 RX MEDLINE; 96213747.
 RA LITJENS T., BROOKS D.A., PETERS C., GIBSON G.J., HOPWOOD J.J.;
 RT "Identification, expression, and biochemical characterization of N-
 RT acetylgalactosamine-4-sulfatase mutations and relationship with N-
 RT clinical phenotype in MPS-VI patients.";
 RL Am. J. Hum. Genet. 58:1127-1134(1996).

[10]
 RN VARIANTS MPS-VI TRP-152 AND GLN-160.
 RX MEDLINE: 94171224.
 RA VOSKOBOEVA E., ISBRANDT D., VON FIGURA K., KRASNOPOLSKAYA X.,
 RA PETERS C.;
 RT "Four novel mutant alleles of the arylsulfatase B gene in two
 RT patients with intermediate form of mucopolysaccharidosis VI
 RT (Maroteaux-Lamy syndrome).";
 RL Hum. Genet. 93:259-264 (1994).
 RN [11]
 RP VARIANT MPS-IV ARG-302.
 RA VILLANI G.R.D., BALZANO N., DI NATALE P.;
 RT "Two novel mutations of the arylsulfatase B gene in two Italian
 RT patients with severe form of mucopolysaccharidosis.";
 RL Hum. Mutat. 11:410-410 (1998).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 4-SULFATE GROUPS OF THE
 CC N-ACETYL-D-GALACTOSAMINE 4-SULFATE UNITS OF CHONDROITIN SULFATE
 CC AND DERMATAN SULFATE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.
 CC DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
 CC SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
 CC THIS DISEASE IS CHARACTERIZED BY THE ACCUMULATION OF DERMATAN
 CC SULFATE IN LYSOSOMES. CLINICAL FEATURES CAN INCLUDE ABNORMAL
 CC GROWTH, SHORT STATURE, STIFF JOINTS, SKELETAL MALFORMATIONS,
 CC CORNEAL CLOUDING, HEPATOSPLENOMEGALY, AND CARDIAC ABNORMALITIES.
 CC A WIDE VARIATION IN CLINICAL SEVERITY IS OBSERVED.
 CC -1- DISEASE: MULTIPLE SULFATASE DEFICIENCY (MSD) IS A DISORDER THAT
 CC COMBINES FEATURES OF METACHROMATIC LEUKODYSTROPHY AND OF
 CC MUCOPOLYSACCHARIDOSIS; IT IS CHARACTERIZED BY A DECREASED
 CC ACTIVITY OF ALL KNOWN SULFATASES. IT SEEMS TO BE CAUSED FROM THE
 CC LACK OF POST-TRANSLATIONAL MODIFICATION OF A CYSTEINE INTO 2-
 CC AMINO-3-OXOPROPIONIC ACID.
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J05225; AAA51784.1; -;
 DR EMBL: M32373; AAA51779.1; -;
 DR EMBL: X72735; CAA51272.1; -;
 DR EMBL: X72736; CAA51272.1; JOINED.
 DR EMBL: X72737; CAA51272.1; JOINED.
 DR EMBL: X72738; CAA51272.1; JOINED.
 DR EMBL: X72739; CAA51272.1; JOINED.
 DR EMBL: X72740; CAA51272.1; JOINED.
 DR EMBL: X72741; CAA51272.1; JOINED.
 DR EMBL: X72742; CAA51272.1; JOINED.
 DR EMBL: S57777; AAB19988.1; -;
 DR PIR: A35078; A35078.
 DR PDB: 1FS0; 04-FEB-98.
 DR MIM: 253200; -;
 DR MIM: 272200; -;
 DR PROSITE: PS00523; SULFATASE_1; 1.
 DR PROSITE: PS00149; SULFATASE_2; 1.
 DR PFAM: PF00884; Sulfatase; 1.
 KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis;
 KW Disease mutation; Polymorphism; 3D-structure.
 FT SIGNAL 1 36 (POTENTIAL).
 FT CHAIN 37 533 ARYL-SULFATASE B.
 FT ACT_SITE 147 147 POTENTIAL.
 FT MOD_RES 91 91 2-AMINO-3-OXOPROPIONIC ACID.
 FT DISULFID 117 521
 FT DISULFID 121 155
 FT DISULFID 181 192
 FT DISULFID 405 447
 FT CARBOHYD 188 188
 FT CARBOHYD 279 279
 FT POTENTIAL.

FT	CARBOHYD	291	291	PROBABLE.
FT	CARBOHYD	366	366	
FT	CARBOHYD	426	426	POTENTIAL.
FT	CARBOHYD	458	458	T -> M (IN MPS-VI; MILD FORM).
FT	VARIANT	92	92	/FTID-VAR_007294.
FT	VARIANT	95	95	R -> Q (IN MPS-VI; MILD/SEVERE FORM).
FT	VARIANT	117	117	/FTID-VAR_007295.
FT	VARIANT	117	117	C -> R (IN MPS-VI; SEVERE FORM).
FT	VARIANT	137	137	/FTID-VAR_007296.
FT	VARIANT	137	137	G -> V (IN MPS-VI; INTERMEDIATE FORM).
FT	VARIANT	152	152	/FTID-VAR_007297.
FT	VARIANT	160	160	R -> W (IN MPS-VI; INTERMEDIATE FORM).
FT	VARIANT	160	160	/FTID-VAR_007298.
FT	VARIANT	210	210	R -> Q (IN MPS-VI; INTERMEDIATE FORM).
FT	VARIANT	210	210	/FTID-VAR_007299.
FT	VARIANT	236	236	Y -> C (IN MPS-VI; MILD/INTERMEDIATE).
FT	VARIANT	236	236	/FTID-VAR_007300.
FT	VARIANT	302	302	L -> P (IN MPS-VI; MILD FORM).
FT	VARIANT	302	302	G -> R (IN MPS-VI; SEVERE FORM).
FT	VARIANT	376	376	/FTID-VAR_007301.
FT	VARIANT	393	393	V -> M.
FT	VARIANT	405	405	/FTID-VAR_007303.
FT	VARIANT	405	405	H -> P (IN MPS-VI; MILD/SEVERE FORM).
FT	VARIANT	498	498	/FTID-VAR_007304.
FT	VARIANT	498	498	C -> Y (IN MPS-VI; MILD FORM).
FT	VARIANT	498	498	/FTID-VAR_007305.
FT	VARIANT	498	498	L -> P (IN MPS-VI; MILD/SEVERE FORM).
FT	VARIANT	533	533	/FTID-VAR_007306.
FT	CONFLICT	533	533	V -> M (IN REF. 3).
SQ	SEQUENCE	533	533	AA; 59687 MW; 986B4C42 CRC32;

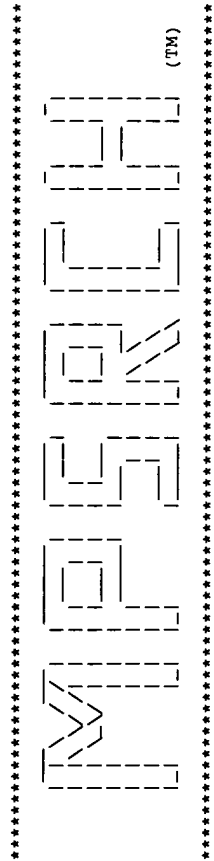
Query Match 4.7%; Score 190; DB 1; Length 533;
 Best Local Similarity 27.1%; Pred. No. 2.04e-15;
 Matches 69; Conservative 70; Mismatches 95; Indels 21; Gaps 17;

Db	9	LRPGGPRRL-LPVLPLLLLLLPPGSGAGARPHLVFLADLGNWVGFGHSR- 66
Qy	1	MPPTGRGLLGLGLVSSVCVAL-GSETQANSTADLNVLIIIVDDL-RPSLGCYGDKL 58
Db	67	IRTPHDLAAGVLLDNYTQ-PLCTPSRSLTLTGRTGL-QHIIWPCQSCVPL 124
Qy	59	VRSPNIDQLASHLLFQNAFAQAQVCAVSRVSLTGRPTDRLYDFNSVWRVHAG--NF 116
Db	125	DEKLLPQLKEAGYTHMVCK-WHLGMYRKECLTRGRFTYCYLLGSDYSHERCTL 183
Qy	117	S-TI-POYFKENGYVTMSVGVKVFHGLSSNHTDDSPYSW-SFPYPHPSSEKYENTKTCRG 173
Db	184	IDA-LNVTR-CALDFRDGEVATGYKNMYSTNIFTKRAIALITN-HPPEKPLFLYLALQS 240
Qy	174	PDGELHANLCPVDVLD--VPEG--TL-PDKQSTEQAIQLLEKMKTSASPFPLAVGYHK 227
Db	241	VHEPLQVPEYLKPY 255
Qy	228	PHIPFRYPKEFKLY 242

RESULT 15
 ID ARS_HEMPU STANDARD; PRT; 551 AA.
 AC P14000;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ARYL-SULFATASE PRECURSOR (EC 3.1.6.1) (ARYL-SULFATE SULPHOHYDROLASE)
 DE (ARS).
 OS Hemichentrotus pulcherrimus (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Hemichentrotus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=PLUTEUS;

RX	MEDLINE; 89030699.
RA	SASAKI H., YAMADA K., AKASAKA H., SUZUKI K., SAITO A., SATO M., SHIMADA H.;
RA	"cDNA cloning, nucleotide sequence and expression of the gene for arylsulphatase in the sea urchin (<i>Hemicentrotus pulcherrimus</i>) embryo."; Eur. J. Biochem. 177:9-13(1988).
RT	[2]
RL	SEQUENCE FROM N.A.
RP	MEDLINE; 90092130.
RX	YAMADA K., AKASAKA K., SHIMADA H.;
RA	"Structure of sea-urchin arylsulphatase gene.";
RT	Eur. J. Biochem. 186:405-410(1989).
CC	-1- FUNCTION: MAY BE A STRUCTURAL COMPONENT OF THE EXTRACELLULAR MATRICES INVOLVED IN CELL MOVEMENT DURING MORPHOGENESIS.
CC	-1- CATALYTIC ACTIVITY: A PHENOL SULFATE + H(2)O -> A PHENOL + SULFATE..
CC	-1- SUBCELLULAR LOCATION: IN BOTH THE CYTOPLASM AND THE EXTRACELLULAR MATRICES.
CC	-1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
DR	EMBL; X17015; CAA34881.1; .
DR	PIR; S01793; S01793.
DR	PIR; S07089; S07089.
DR	HSP; P15289; IADK.
DR	PROSITE; PS00523; SULFATASE_1; 1.
DR	PROSITE; PS00149; SULFATASE_2; 1.
DR	PFAM; PF00884; Sulfatase; 1.
KW	Hydrolase; Signal; Glycoprotein; Extracellular matrix.
FT	SIGNAL 1 20
FT	CHAIN 21 551 ARYL SULFATASE.
FT	MOD_RES 21 21 BLOCKED.
FT	MOD_RES 100 100 2-AMINO-3-OXOPROPIONIC ACID (BY SIMILARITY).
FT	ACT_SITE 158 158 POTENTIAL.
FT	CARBOHYD 164 164 POTENTIAL.
FT	CARBOHYD 213 213 POTENTIAL.
FT	CARBOHYD 296 296 POTENTIAL.
SQ	SEQUENCE 551 AA; 60952 MW; 0C7B7505 CRC32;
	Query Match 4.5%; Score 184; DB 1; Length 551;
	Best Local Similarity 32.8%; Pred. No. 2.73e-14;
	Matches 39; Conservative 28; Mismatches 45; Indels 7; Gaps 6;
Dn	53 NVLLVLVADHMGSGDLTSYGHTPTQAGFIDKMAEGLRFTNGYVGDAVCTPSRSAIMTGRL 112 ::: : : : : : : : : : : : : : : : : : : : : : : : :
Qy	38 NVLLIIIVDLLRPS-LGCGYDGKLVRSPNIDLQASHSLFLQNFAQAQAVCAPSRVSFLTGR 96 ::: : : : : : : : : : : : : : : : : : : : : :
Dn	113 PVRIGTGETRVFLPWTKTGLPKSELTIAEAMKEPATGVGMVK-WHLGINENSSTDGA 170 : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy	97 PDTRTRYL-DFNYSY--W-RVHAGNFS-TIPQYFKENGVTMTSVKGVFHPGISNNHTDDSP 150 : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: Wed Apr 19 21:15:54 2000
Job time : 395 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 20:58:22 2000; MasPar time 35.76 Seconds
Tabular output not generated. 725.469 Million cell updates/sec

Title: >US-09-249-003-2
Description: (1-550) from US09249003.pep
Perfect Score: 4069
Sequence: 1 MPPRTVGRGLLWGLVLSV.....QDHNNYNDQGGDLFQLLMP 550

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 51.177; Variance 98.507; scale 0.520

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	No.	Score	Query	Match	Length	ID	Pred. No.
1	4069	100.0	550	1	KJHUID	iduronate-2-sulfatase	0.00e+00
2	3129	76.9	563	2	A47153	iduronate-2-sulfatase	0.00e+00
3	1300	31.9	212	2	PN0565	iduronate-2-sulfatase	1.72e-242
4	282	6.9	497	2	G65169	probable sulfatase (E	1.02e-29
5	239	5.9	589	2	I37187	arylsulfatase E (EC 3	7.72e-22
6	234	5.2	593	2	I37186	arylsulfatase D (EC 3	6.07e-21
7	211	5.2	522	1	KJHUG6	N-acetylglucosamine	6.84e-17
8	211	5.2	567	2	A37362	arylsulfatase (EC 3.1	6.84e-17
9	213	5.2	583	1	KJHUAC	steryl-sulfatase (EC	3.07e-17
10	198	4.9	506	2	A54190	cerebroside-sulfatase	1.17e-14
11	195	4.8	507	1	KJHUAB	cerebroside-sulfatase	3.80e-14
12	190	4.7	533	1	KJHUAH	N-acetylglucosamine	2.65e-13
13	190	4.7	551	2	S07089	arylsulfatase (EC 3.1	2.65e-13
14	184	4.5	551	2	S01793	arylsulfatase (EC 3.1	2.65e-12
15	182	4.5	970	2	E70533	probable sulfatase (E	5.70e-12
16	179	4.4	465	2	F70837	probable sulfatase (E	1.78e-11
17	172	4.2	535	1	A44475	N-acetylglucosamine	2.47e-10
18	153	3.8	313	1	Q0ECIL	conserved hypothetical	2.52e-07
19	153	3.8	551	2	S30691	arylsulfatase (EC 3.1	2.52e-07
20	136	3.3	787	2	B06433	probable sulfatase (E	8.97e-05
21	133	3.3	787	2	B70535	probable sulfatase (E	2.44e-04
22	126	3.1	433	1	A34227	transcobalamin I prec	2.39e-04
23	126	3.1	571	2	E64903	arylsulfatase homolog	2.39e-03

24	126	3.1	919	2	S37786	hypothetical protein	2.39e-03
25	124	3.0	464	2	B35159	arylsulfatase (EC 3.1	4.53e-03
26	120	2.9	190	1	VCBVC	coat protein - cowpea	1.60e-02
27	117	2.9	533	2	S69336	arylsulfatase (EC 3.1	4.04e-02
28	115	2.8	473	2	I54210	arylsulfatase (EC 3.1	7.43e-02
29	114	2.8	540	1	OXECLD	L-aspartate oxidase (1.00e-01
30	112	2.8	585	1	F64159	hypothetical protein	1.83e-01
31	111	2.7	708	2	B72619	probable NADH dehydro	2.46e-01
32	109	2.7	709	2	T16584	hypothetical protein	4.42e-01
33	105	2.6	44	2	B38075	N-acetylglactosamine	1.40e+00
34	107	2.6	479	2	C75099	hypothetical protein	7.90e-01
35	107	2.6	552	1	KJHUGU	N-acetylglucosamine-6	7.90e-01
36	105	2.6	647	2	C71534	probable transglycola	1.40e+00
37	106	2.6	667	2	C70719	probable carbamoyl-ph	1.05e+00
38	106	2.6	815	2	T15749	hypothetical protein	1.05e+00
39	107	2.6	1034	2	S36758	mgll protein - mouse	7.90e-01
40	107	2.6	4563	1	LPHUB	apolipoprotein B-100	7.90e-01
41	102	2.5	245	2	B33956	tyrocidine synthetase	3.24e+00
42	103	2.5	440	2	S73396	hypothetical protein	2.45e+00
43	103	2.5	467	2	E70976	hypothetical protein	2.45e+00
44	102	2.5	518	1	STCEEC	glutamate--cysteine 1	3.24e+00
45	102	2.5	586	1	C64988	hypothetical 67.3 kD	3.24e+00

ALIGNMENTS

KJHUID #type complete
iduronate-2-sulfatase (EC 3.1.1.6.13), precursor - human
chondroitin-sulfatase; L-iduronate 2-sulfate 2-sulfohydrolase
#formal name Homo sapiens #common name man
28-Mar-1991 #sequence_revision 27-Oct-1995 #text_change
24-Sep-1999
A47535; A36483
A47535
Wilson, P.J.; Meaney, C.A.; Hopwood, J.J.; Morris, C.P.
Genomics (1993) 17:773-775
Sequence of the human iduronate 2-sulfatase (IDS) gene.
#cross-references MUID:94063929
#accession A47535
#status translation not shown; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-550 #label W12
#cross-references GB:LI3329; NID:g405203; PIDN:AAA16877.1; PID:g405205
A36483
Wilson, P.J.; Morris, C.P.; Anson, D.S.; Occhiodoro, T.;
Bielicki, J.; Clements, P.R.; Hopwood, J.J.
Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8531-8535
Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA
clone and analysis of patient DNA.
#cross-references MUID:91046030
#accession A36483
#molecule_type mRNA
#cross-references GB:M58342; GB:M38371; NID:g184561; PIDN:AAA63197.1; PID:g184562
#note parts of this sequence, including the amino end of the
a mature protein, were determined by protein sequencing
#note a form is described with a proteolytic cleavage
releasing residue 456 as the amino terminal of the
cleaved chain
GDB:IDS; SIDS
#cross-references GDB:120521; OMIM:309900
#map_position Xq28-Xq28
#introns 3571; 80/3; 140/1; 169/3; 236/3; 293/3; 336/1; 394/1
#note defects in this gene can cause mucopolysaccharidosis type II,
Hunter disease
#description hydrolyzes iduronate-2-sulfate units in dermatan sulfate and
heparan sulfate to release sulfate
#superfamily animal sulfatase
CLASSIFICATION
glycoprotein; Hunter disease; lysosomal storage disease;


```
lysosome; sulfuric ester hydrolase

FEATURE
1-25      #domain signal sequence #status predicted #label SIG\
26-33     #domain propeptide #status predicted #label PRO\
34-550    #product iduronate-2-sulfatase #status predicted #label
          MAT\
31,115,144,246,280,
325,513,537  #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
84          #modified_site 3-oxoalanine (Cys) #status predicted
SUMMARY     #length 550 #molecular-weight 61872 #checksum 5247
            100.0%; Score 4069; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPPRTGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLYR 60
QY 1 MPPRTGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLYR 60
Db 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPRDTRTRYDFNSYWRVHAGNFSTIP 120
QY 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPRDTRTRYDFNSYWRVHAGNFSTIP 120
Db 121 QYFKENGVTMSVGVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRGPDELHA 180
QY 121 QYFKENGVTMSVGVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRGPDELHA 180
Db 181 NLLCPVDVLDVPEGLTPDKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQK 240
QY 181 NLLCPVDVLDVPEGLTPDKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQK 240
Db 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIROEDVQALNISVPYGPPIPVDFQKIRQSY 300
QY 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIROEDVQALNISVPYGPPIPVDFQKIRQSY 300
Db 301 FASVSYLDTQVGRLLSALDDQLANSTTIAFTSDHGWLGEHGEWAKYSNFDVATHVPLI 360
QY 301 FASVSYLDTQVGRLLSALDDQLANSTTIAFTSDHGWLGEHGEWAKYSNFDVATHVPLI 360
Db 361 FYVPGRTASLPENCKGLFPYLDPPDSASOLMEPGQSDMLVELSLFPTLAGLAGLOVPP 420
QY 361 FYVPGRTASLPENCKGLFPYLDPPDSASOLMEPGQSDMLVELSLFPTLAGLAGLOVPP 420
Db 421 RCPVPFHVLCREGKLLKHFRRDLEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480
QY 421 RCPVPFHVLCREGKLLKHFRRDLEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480
Db 481 SLKDIKIMGYSIRTDYRYTVWVGFPNDFELANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
QY 481 SLKDIKIMGYSIRTDYRYTVWVGFPNDFELANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
Db 541 GGDLFQLLMP 550
QY 541 GGDLFQLLMP 550

RESULT 2
ENTRY A47153 #type complete
TITLE iduronate-2-sulfatase (EC 3.1.1.6.13) - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
22-Jun-1999
ACCESSIONS A47153
REFERENCE
#authors Daniele, A.; Faust, C.J.; Herman, G.E.; Di Natale, P.;
Ballabio, A.
#journal Genomics (1993) 16:755-757
#title Cloning and characterization of the cDNA for the murine
iduronate sulfatase gene.
#cross-references MUID:93315172
#accession A47153
#status preliminary
```

```
##molecule_type mRNA
##residues 1-563 ##label DAN
##cross-references GB:L07921; NID:g349713; PIDN:AAA37880.1; PID:g349714
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS sulfuric ester hydrolase
FEATURE
99 #modified_site 3-oxoalanine (Cys) #status predicted
SUMMARY #length 563 #molecular-weight 63437 #checksum 8505

Query Match 76.9%; Score 3129; DB 2; Length 563;
Best Local Similarity 79.0%; Pred. No. 0.00e+00;
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps 4;

Db 19 ARAIWQRLSFLLSGFCIALESAAQNSATDALNILLIIVDDLRLPSLCYGDKLYRSPN 78
QY 4 PRTRGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLYRSPN 63
Db 79 IDQLASHSVLFQNAFAQAQAVCAPSRVSFLTGRPRDTRTRYDFNSYWRVHAGNFSTIPQYF 138
QY 64 IDQLASHSVLFQNAFAQAQAVCAPSRVSFLTGRPRDTRTRYDFNSYWRVHAGNFSTIPQYF 123
Db 139 KENGYVTMSVGVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRGPDELHANLL 198
QY 124 KENGYVTMSVGVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRGPDELHANLL 183
Db 199 CPVDVADVPEGLTPDKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQKLYP 258
QY 184 CPVDVLDVPEGLTPDKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQKLYP 243
Db 259 LENITLAPDPEVPGDGLPPVAYNPWMDIROEDVQALNISVPYGPPIPVDFQKIRQSYFAS 318
QY 244 LENITLAPDPEVPGDGLPPVAYNPWMDIROEDVQALNISVPYGPPIPVDFQKIRQSYFAS 303
Db 319 VSYLDTQVGRLLSALDDQLANSTTIAFTSDHGWLGEHGEWAKYSNFDVATHVPLIYV 378
QY 304 VSYLDTQVGRLLSALDDQLANSTTIAFTSDHGWLGEHGEWAKYSNFDVATHVPLIFYV 363
Db 379 PGRTAPLPAAGKLPYLDPPDSASOLMEPGQSDMLVELSLFPTLAGLAGLOVPP 438
QY 364 PGRTASLPENCKGLFPYLDPPDSASOLMEPGQSDMLVELSLFPTLAGLAGLOVPPRCP 423
Db 439 SLLFMLSFAEKAR-IPRSIC-SSMTWKRQTCCLVPRELIAYSQYPRPADPQWNSDKP 496
QY 424 VPSFHVLCREGKLLKHFRRDLE-EDPYLPGNPRELIAYSQYPRPSDIPQWNSDKPSL 482
Db 497 NDIRSHWDIYTH-VDYRYTVWVGFPDSEFLANFSDIHAGELYFVDSPLQDHNMYNDSQHG 555
QY 483 KDIKIMGYSIRTDYRYTVWVGFPNDFELANFSDIHAGELYFVDSPLQDHNMYNDSQGG 542
Db 556 GLLHSLRP 563
QY 543 DLFQLLMP 550

RESULT 3
ENTRY PNO565 #type fragment
TITLE iduronate-2-sulfatase (EC 3.1.1.6.13) (clone MTA13) - mouse
ALTERNATE_NAMES (fragment)
ORGANISM iduronate sulfatase
#formal_name Mus musculus #common_name house mouse
DATE 05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change
17-Mar-1999
ACCESSIONS PNO565
REFERENCE
#authors Daniele, A.; Russo, T.; Ballabio, A.; Di Natale, P.;
#journal Biochem. Biophys. Res. Commun. (1993) 194:1030-1037
#title The mouse iduronate sulfatase gene: Identification of a novel
transcript.
#cross-references MUID:93356764
#accession PNO565
##molecule_type mRNA
##residues 1-212 ##label DAN
##note the authors translated the codon CAT for residue 23 as
```

QY 97 PDTIRLIDENSIWRVHAGNFSTLPQYFKENGYVTMSVGKVFH-PGISSNHTDDSPYSWSF 155

4

RESULTS

```

ENTRY          A54190      #type complete
TITLE          cerebroside-sulfatase (EC 3.1.6.8) precursor - mouse
ALTERNATE_NAMES
ORGANISM       arylsulfatase A
#formal_name Mus musculus #common_name house mouse
DATE           18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
29-Sep-1999
ACCESSIONS     A54190; S34293; A38075; S34294
REFERENCE
#authors       Kreyling, J.; Polten, A.; Hess, B.; von Figura, K.; Menz, K.;
Steiner, F.; Gieselmann, V.
#journal       Genomics (1994) 19:249-256
#title        Structure of the mouse arylsulfatase A gene and cDNA.
#cross-references EMBL:94245194
#accession     A54190
#status        preliminary
#molecule_type mRNA
#residues      1-506 #label KRE
#cross-references GB:X73230; NID:g312271; PIDN:CAA51702.1; PID:g312272
REFERENCE
#authors       Kreyling, J.; Polten, A.; Hess, B.; Menz, K.; Steiner, F.;
Gieselmann, V.
#submission    submitted to the EMBL Data Library, June 1993
#accession     S34293
#status        preliminary
#molecule_type DNA
#residues      1-506 #label KR2
#cross-references EMBL:X73231; NID:g312268; PIDN:CAA51703.1;
PID:g312269
REFERENCE
#authors       Grompe, M.; Pieretti, M.; Caskey, C.T.; Ballabio, A.
#journal       Genomics (1992) 12:755-760
#title        The sulfatase gene family: cross-species PCR cloning using
the MOPAC technique.
#cross-references MUID:92241876
#accession     A38075
#molecule_type mRNA
#residues      27-72 #label GRO
#cross-references GB:M82876; NID:gl92027; PIDN:AAA37260.1; PID:gl92028
#note          sequence extracted from NCBI backbone (NCBIN:98744,
NCBIP:98749)
#note          the protein sequence from Fig. 3 includes conserved
regions used in the primer
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS        glycoprotein; sulfuric ester hydrolase
FEATURE
1-17            #domain signal sequence #status predicted #label SIG\
18-506          MAT\
68              #modified_site 3-oxoalanine (Cys) #status predicted\
157,183,349     #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY          #length 506 #molecular-weight 53776 #checksum 7525
Query Match      4.9%; Score 198; DB 2; Length 506;
Best Local Similarity 42.5%; Pred. No. 1.17e-14;
Matches 31; Conservative 16; Mismatches 25; Indels 1; Gaps 1;
Db 15 STASPPNLLIFADDLGVGLSGVHSPSTPNLDLAEGLRFTDFVVPVSLCTPSRAA 74
||: :|:||||: ||| ||| :||:||||: ||: :|:||||:
QY 32 STTDALNVLIIIVDLR-PSLGCYGDKLVRSPNIDQLASHSLLFQNAFAQAQVACPSRVS 90
||:||||| :|: ||| ||| :|: ||| ||| :|: ||| |||
Db 75 LLTGRLPVRSAMY 87
QY 91 FLTGRRPDTRRLY 103
RESULT 11
ENTRY          KJHUA      #type complete
TITLE          cerebroside-sulfatase (EC 3.1.6.8) precursor - human
ALTERNATE_NAMES arylsulfatase A (ASA); arylsulfatase (EC 3.1.6.1)
(misidentification)
ORGANISM       #formal_name Homo sapiens #common_name man
DATE           21-Nov-1993 #sequence_revision 27-Oct-1995 #text_change

ACCESSIONS     S11031; G02857; A32207; S23932
REFERENCE
#authors       Kreyling, J.; von Figura, K.; Gieselmann, V.
#journal       Eur. J. Biochem. (1990) 191:627-631
#title        Structure of the arylsulfatase A gene.
#cross-references MUID:90361046
#accession     S11031
#molecule_type DNA
#residues      1-507 #label KRE
#cross-references EMBL:X52150; NID:g28859; PIDN:CAA36398.1; PID:g28860
REFERENCE
#authors       Adams, M.D.; Kerlavage, A.R.; Fuldner, R.A.; Phillips, C.A.;
Venter, J.C.
#submission    submitted to the EMBL Data Library, June 1996
#accession     G02857
#status        translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      'MS', 1-507 #label ADA
#cross-references EMBL:U62317; NID:gl399959; PID:gl3
#note          an incorrect initiation codon was used
REFERENCE
#authors       Stein, C.; Gieselmann, V.; Kreyling, J.; Schmidt, B.;
Pohlmann, K.; Waheed, A.; Meyer, H.E.; O'Brien, J.S.; von
Figura, K.
#journal       J. Biol. Chem. (1989) 264:1252-1259
#title        Cloning and expression of human arylsulfatase A.
#cross-references MUID:89093115
#accession     A32207
#molecule_type mRNA
#residues      1-358, 'RPPAAGHRQPSAVSLLLPVLPRGPGWGCADMKVQGSLLHP',
402-507 #label STE
#cross-references GB:X52151; GB:J04442; GB:J04593
#note          parts of this sequence, including the amino end of the
mature protein, were determined by protein sequencing
REFERENCE
#authors       Fujii, T.; Kobayashi, T.; Honke, K.; Gasa, S.; Ishikawa, M.;
Shimizu, T.; Makita, A.
#journal       Blochim. Biophys. Acta (1992) 1122:93-98
#title        Protolytic processing of human lysosomal arylsulfatase A.
#cross-references MUID:92338230
#accession     S23932
#molecule_type protein
#residues      20-29, 31-33; 434-479 #label FUJ
#experimental_source placenta
REFERENCE
#authors       Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K.
#journal       Cell (1995) 82:271-278
#title        A novel amino acid modification in sulfatases that is
defective in multiple sulfatase deficiency.
#cross-references MUID:95354208
#contents      annotation; identification of 3-oxoalanine,
2-amino-3-oxopropanoic acid
GENETICS
#gene          GDB:ARSA
#cross-references GDB:119007; OMIM:250100
#map_position  22q13.31-22qter
#introns       73/2; 153/3; 226/3; 283/2; 325/1; 367/3; 402/1
#note          defects in this gene can cause metachromatic leukodystrophy
FUNCTION
#description    hydrolyzes cerebroside 3-sulfate to release sulfate; can also
hydrolyze galactose-3-sulfate, ascorbate 2-sulfate, and
many phenol sulfates
#superfamily animal sulfatase
KEYWORDS        glycoprotein; lysosomal storage disease; lysosome; sulfuric
ester hydrolase
FEATURE
1-19           #domain signal sequence #status predicted #label SIG\
20-507         #product cerebroside-sulfatase #status predicted #label
MAT\
20-444,445-507 #product cerebroside-sulfatase component b #status
predicted #label MCB\
20-444,448-507 #product cerebroside-sulfatase component c #status

```

```

#procedure-type protein
##residues 41-55;424-425,'X',427-454;466-483 ##label KOB

```


DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Sep-1999

ACCESSIONS E70533

REFERENCE A70500

#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544

#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references MUID:98295987

#accession E70533

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

##residues 1-970 ##label COL

##cross-references GB:Z96070; GB:AL123456; NID:g3261791; PID:e320915; PID:g2181956

##experimental_source strain H37Rv

GENETICS

#gene atsB

KEYWORDS sulfuric ester hydrolase

FEATURE 258

SUMMARY #length 970 #molecular-weight 105679 #status predicted #modified_site 3-oxoalanine (Cys) #checksum 7320

Query Match 4.5%; Score 182; DB 2; Length 970;

Best Local Similarity 30.0%; Pred. No. 5.70e-12;

Matches 39; Conservative 39; Mismatches 40; Indels 12; Gaps 9;

Db 210 DAPNVLIVLIDDAGFGPDTEGGA-IRPTLSRLAQNGLIY-NRFHVAVCSPTRAALLT 267

Qy 35 DALNVLLIIIVDDLRL-PSLGCYGDKLVRSPNIDQLASHSLLFQNAFAQAQVAFPSRVSF 93

Db 268 GRNHRVGFSGVCEFPVPGYSAVRPSCAALPRILRDNGYVTGAFGR-WHLTPDNVQG 326

Qy 94 GR--RPDTTRIYDF-NSY--WR-VHAGNFSTIPQTKENGIVTMSVGKVFHPGISSNHT 146

Db 327 AAGPFDNWPL 336

Qy 147 DDSPY-SWSF 155

Search completed: Wed Apr 19 21:09:02 2000

Job time : 640 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 20:52:10 2000; MasPar time 16.52 Seconds
Tabular output not generated. 788.664 Million cell updates/sec

Title: >US-09-249-003-2

Description: (1-550) from US09249003.pep

Perfect Score: 4069
Sequence: 1 MPPTPTGRGLLWGLVLSSV.....QDHNMYNDSQGDLFQLLMP 550

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:genesep

Statistics: Mean 37.057; Variance 165.341; scale 0.224

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4069	100.0	550	1	Human iduronate-2-sulf	0.00e+00
2	4069	100.0	550	1	Human IDS.	0.00e+00
3	173	4.3	406	1	Human secreted protein	1.44e-04
4	176	4.3	502	1	Human sulphamidase pro	8.28e-05
5	172	4.2	556	1	Sequence of murine bon	1.73e-04
6	164	4.0	86	1	Human 5' ESR secreted	7.40e-04
7	136	3.3	25	1	Human iduronate-2-sulf	1.05e-01
8	136	3.3	25	1	Human IDS peptide frag	1.05e-01
9	109	2.7	226	1	Flea cysteine protease	9.52e+00
10	109	2.7	327	1	Flea cysteine protease	9.52e+00
11	109	2.7	327	1	Flea cysteine protease	9.52e+00
12	110	2.7	1671	1	Signal peptide and fir	8.10e+00
13	110	2.7	2721	1	Sequence of N-terminal	8.10e+00
14	107	2.6	351	1	Caffeine demethylase.	1.31e+01
15	107	2.6	4536	1	Amino acid sequence of	1.31e+01
16	107	2.6	4536	1	Apolipoprotein B-100.	1.31e+01
17	102	2.5	197	1	Canine Fc-epsilon-rece	2.89e+01
18	102	2.5	197	1	Canine Fc-epsilon-rece	2.89e+01
19	102	2.5	197	1	Canine Fc-epsilon-R (1	2.89e+01
20	102	2.5	253	1	Canine Fc-epsilon-rece	2.89e+01
21	101	2.5	353	1	Salmonella enterica O	3.37e+01
22	100	2.5	710	1	Eph-related PTK Cek7.	3.94e+01
23	100	2.5	722	1	Eph-related PTK Cek7.	3.94e+01

24 100 2.5 744 1 R75713 Eph-related PTK Cek7+.

25 99 2.4 199 1 W81111 Canine Fc-epsilon-rece

26 96 2.4 1074 1 R24102 Marek's disease virus

27 97 2.4 1301 1 R28304 Acid alpha-amylase enz

28 96 2.4 1676 1 R77604 Pro-C5 polypeptide.

29 95 2.3 16 1 W69787 Human iduronate-2-sulf

30 95 2.3 16 1 W47299 Human IDS peptide frag

31 93 2.3 104 1 R80087 Human derived light ch

32 93 2.3 104 1 W95485 Human-derived RN3 phag

33 92 2.3 258 1 W08487 C6 human sfv antibody

34 92 2.3 262 1 W97890 C6.5/218 single chain

35 94 2.3 614 1 W38320 Specific co-activator

36 92 2.3 817 1 R47201 DPM2 mannosyl transfer

37 93 2.3 862 1 W31655 Pacific yew taxadiene

38 90 2.2 252 1 W72049 HSV-2 strain SB5 Conti

39 90 2.2 331 1 R64769 HSV-2 ISP35 protein.

40 90 2.2 455 1 R39344 Epic protein.

41 90 2.2 490 1 R41663 Paired basic amino aci

42 90 2.2 636 1 R64767 HSV-2 protease, ICP35.

43 90 2.2 638 1 R71017 HSV-2 protease.

44 90 2.2 642 1 W72124 HSV-2 strain SB5 Conti

45 91 2.2 849 1 R86595 N-tera 2D1 autotaxin p

ALIGNMENTS

RESULT 1

ID W69786 standard; Protein; 550 AA.

AC W69786;

DT 03-NOV-1998 (first entry)

DE Human iduronate-2-sulfatase.

KW Human; iduronate-2-sulfatase; IDS; liver; glycosylated; heparin sulphate;

KW dermatan sulphate; lysosome; Hunter syndrome; severe mental retardation;

KW skeletal deformity; stiff joint.

OS Homo sapiens.

PN US5798239-A.

PD 25-AUG-1998.

PF 07-JUN-1995; 484494.

PR 17-DEC-1992; US-991973.

PR 12-NOV-1991; US-790362.

PR 28-NOV-1994; US-345212.

PR 07-JUN-1995; US-484494.

PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.

PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,

PI Occhiodoro T, Wilson PJ;

DR WPI; 98-480382/41.

DR N-PSDB: V52836.

PT Production of glycosylated iduronate-2-sulphatase enzyme (IDS) -

PT comprises culturing a host cell containing a nucleic acid encoding

PT the enzymatically active iduronate-2-sulphatase polypeptide

PS Example 1; Fig 1; 53pp; English.

CC A method has been developed for the production of glycosylated

CC iduronate-2-sulphatase enzyme (IDS). The method comprises culturing

CC a host cell containing a nucleic acid encoding the enzymatically

CC active IDS polypeptide where the host cell glycosylates the polypeptide

CC to a greater degree than a native IDS polypeptide expressed by a

CC natural human liver cell. The present sequence represents human

CC IDS which is used in an example from the present invention. The

CC recombinant IDS is used to treat IDS deficiency where heparin sulphate

CC and dermatan sulphate accumulates in lysosomes resulting in Hunter

CC syndrome which is manifested by e.g. severe mental retardation,

CC skeletal deformities and stiff joints. The recombinant IDS possesses

CC inter alia an improved half life inter alia and improved uptake

CC properties in comparison to the naturally glycosylated molecule.

SQ Sequence 550 AA;

Query Match 100.0%; Score 4069; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 550; Conservative 0; Indels 0; Gaps 0;

Db 1 MPPTPTGRGLLWGLVLSSVCAVGSETQANSTTDALNVLIIIVDDLRPSLCGYGDKLVR 60

QY 1 MPPTPTGRGLLWGLVLSSVCAVGSETQANSTTDALNVLIIIVDDLRPSLCGYGDKLVR 60

Db 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120
 QY 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120
 Db 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYHPSPSEKYENTKTCRGPDELHA 180
 QY 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYHPSPSEKYENTKTCRGPDELHA 180
 Db 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQOK 240
 QY 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQOK 240
 Db 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNLSVPYGPVDFQFKIROSY 300
 QY 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNLSVPYGPVDFQFKIROSY 300
 Db 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360
 QY 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360
 Db 361 FYVPGRTASLPEAGEKLPYLDPPDSASQLMPEGRQSDMLVELYSLPPTLAGLAGLOVPP 420
 QY 361 FYVPGRTASLPEAGEKLPYLDPPDSASQLMPEGRQSDMLVELYSLPPTLAGLAGLOVPP 420
 Db 421 RCPVPSHVELCREGKLLKHFRFRDLEEDPYLPGNPRELIAYSOYPRPSDIPQWNSDKP 480
 QY 421 RCPVPSHVELCREGKLLKHFRFRDLEEDPYLPGNPRELIAYSOYPRPSDIPQWNSDKP 480
 Db 481 SLKDIKIMGYSIRTIDYRYTVWVGFNDEFANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
 QY 481 SLKDIKIMGYSIRTIDYRYTVWVGFNDEFANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
 Db 541 GGDLFQLLMP 550
 QY 541 GGDLFQLLMP 550

RESULT 2

ID W47298 standard; Protein; 550 AA.
 AC W47298;
 DT 01-JUN-1998 (first entry)
 DE Human IDS.
 KW Human; iduronate 2-sulphatase; IDS; treatment;
 KW Hunter syndrome.
 OS Homo sapiens.
 PN US5728381-A.
 PD 17-MAR-1998.
 PF 07-JUN-1995; 484493.
 PR 17-DEC-1992; US-991973.
 PR 12-NOV-1991; US-790362.
 PR 28-NOV-1994; US-345212.
 PR 07-JUN-1995; US-484493.
 PA (ANSO/) ANSON D S.
 PA (BIEL/) BIELICKI J.
 PA (CLEM/) CLEMENTS P R.
 PA (HOPW/) HOPWOOD J J.
 PA (MORR/) MORRIS C P.
 PA (OCCH/) OCCHIODORO T.
 PA (WILS/) WILSON P J.
 PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,
 PI Occhiodoro T, Wilson PJ;
 DR WPI: 98-206530/18.
 DR N-PSDB; V15698.
 PT Treatment of iduronate 2-sulphatase deficiency - comprises
 PT administering recombinant iduronate 2-sulphatase
 PS Claim 6; Columns 23-28; 53pp; English.
 CC The present sequence is human iduronate 2-sulphatase (IDS).
 CC IDS deficiency can be treated by administering a recombinant human
 CC IDS that is more highly glycosylated than the naturally occurring
 CC enzyme, useful in the treatment of Hunter syndrome. The recombinant
 CC IDS may be administered in 0.5 microg/kg to 20 mg/kg doses. The
 CC administration route is oral, intravenous, intraperitoneal,

CC intramuscular, subcutaneous or intranasal. The recombinant IDS has
 CC better uptake properties and/or a longer half-life in vivo, and is
 CC thus more efficient than naturally glycosylated IDS.
 SQ Sequence 550 AA;
 Query Match 100.0%; Score 4069; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MPPPTGRGLLWGLVLSVVCVALGSETQANSTTDALNVLIIIVDDLRSLGCGDKLVR 60
 QY 1 MPPPTGRGLLWGLVLSVVCVALGSETQANSTTDALNVLIIIVDDLRSLGCGDKLVR 60
 Db 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120
 QY 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120
 Db 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYHPSPSEKYENTKTCRGPDELHA 180
 QY 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYHPSPSEKYENTKTCRGPDELHA 180
 Db 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQOK 240
 QY 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQOK 240
 Db 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNLSVPYGPVDFQFKIROSY 300
 QY 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIRQEDVQALNLSVPYGPVDFQFKIROSY 300
 Db 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360
 QY 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360
 Db 361 FYVPGRTASLPEAGEKLPYLDPPDSASQLMPEGRQSDMLVELYSLPPTLAGLAGLOVPP 420
 QY 361 FYVPGRTASLPEAGEKLPYLDPPDSASQLMPEGRQSDMLVELYSLPPTLAGLAGLOVPP 420
 Db 421 RCPVPSHVELCREGKLLKHFRFRDLEEDPYLPGNPRELIAYSOYPRPSDIPQWNSDKP 480
 QY 421 RCPVPSHVELCREGKLLKHFRFRDLEEDPYLPGNPRELIAYSOYPRPSDIPQWNSDKP 480
 Db 481 SLKDIKIMGYSIRTIDYRYTVWVGFNDEFANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
 QY 481 SLKDIKIMGYSIRTIDYRYTVWVGFNDEFANFSDIHAGELYFVDSPLQDHNMYNDSQ 540
 Db 541 GGDLFQLLMP 550
 QY 541 GGDLFQLLMP 550

RESULT 3

ID Y02687 standard; Protein; 406 AA.
 AC Y02687;
 DT 11-JUN-1999 (first entry)
 DE Human secreted protein encoded by gene 38 clone HSLBP69.
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 PN WO9902546-A1.
 PD 21-JAN-1999.
 PF 07-JUL-1998; U13684.
 PR 12-SEP-1997; US-058785.
 PR 08-JUL-1997; US-051916.
 PR 08-JUL-1997; US-051918.
 PR 08-JUL-1997; US-051919.
 PR 08-JUL-1997; US-051920.
 PR 08-JUL-1997; US-051925.

CC terminal peptide used in an example from the present invention. The
 CC recombinant IDS is used to treat IDS deficiency where heparin sulphate
 CC and dermatan sulphate accumulates in lysosomes resulting in Hunter
 CC syndrome which is manifested by e.g. severe mental retardation,
 CC skeletal deformities and stiff joints. The recombinant IDS possesses
 CC inter alia an improved half life inter alia and improved uptake
 CC properties in comparison to the naturally glycosylated molecule.
 SQ Sequence 25 AA;

Query Match 3.3%; Score 136; DB 1; Length 25;
 Best Local Similarity 84.0%; Pred. No. 1.05e-01;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 TSALNVLIIIVDDLRPSIGDYDDVL 25
 | ||||| ||||| ||||| ||||| |||||
 QY 34 TDALNVLIIIVDDLRPSIGCGDKL 58

RESULT 9

ID W41994 standard; Protein; 226 AA.
 AC W41994;
 DT 02-JUL-1998 (first entry)
 DE Flea cysteine protease SEQ ID NO:8.
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
 KW Immunoglobulin protease; larvae; host animal.
 OS Siphonaptera.
 PN WO9740058-A1.
 PD 30-OCT-1997.
 PF 24-APR-1997; U06121.
 PR 04-APR-1997; US-042945.
 PR 24-APR-1996; US-639075.
 PR 15-NOV-1996; US-749699.
 PA (HESK-) HESKA CORP.
 PI Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE,
 PI Silver G, Stiegler GL;
 DR WPI: 98-076762/07.
 DR N-PSDB; V04624.
 PT New flea protease genes and proteins - used in vaccine compositions
 PT for the prophylaxis and treatment of flea infestation, especially in
 PT cats or dogs
 PS Claim 2; Page 174-175; 318pp; English.
 CC The present sequence represents a novel flea cysteine protease. The
 CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
 CC protein, as well as the DNA encoding the protein, may all be used in
 CC therapeutic compositions to reduce flea protease activity (especially
 CC immunoglobulin protease) and so reduce flea infestation, especially in
 CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
 CC fleas which comprises anti-protease antibodies produced by a host animal
 CC in response to administration of the protein. Therapeutic compositions
 CC may further comprise a compound that reduces haematophagous ectoparasite
 CC burden by a method other than by reducing flea immunoglobulin protease
 CC activity. The novel flea DNA encoding the protein can also be used to
 CC produce recombinant protein, and fragments of it are used as probes and
 CC primers for identification and isolation of related sequences, also as
 CC antisense, triplex-forming agents and ribozymes for inhibition of the
 CC synthesis of the protein. Ab are also useful for screening expression
 CC libraries, to purify the protein and to target cytotoxins to fleas.
 SQ Sequence 226 AA;

Query Match 2.7%; Score 109; DB 1; Length 226;
 Best Local Similarity 25.3%; Pred. No. 9.52e+00;
 Matches 22; Conservative 26; Mismatches 35; Indels 4; Gaps 4;
 Db 119 RNDVDMKVALKHGPISTVAIDASHKTSFYSNGVYQKCGNRGOLDHVLVVG-YGE 177
 | : || : : : : || : : : : | : : : : | : : : : | : : : : | : : : :
 QY 273 REDVQALNISV-PYGPIPVDFQKIRQS-YFASVSYLDTQVGRLLSALDQLANSTIIA 330
 : | : : : : | : : : : | : : : : | : : : : | : : : :
 Db 178 INSEPYWLKPNQGLMGNVIFDVAKN 204
 : | : : : : | : : : : | : : : : | : : : : | : : : :
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356

RESULT 10

Query Match 2.7%; Score 109; DB 1; Length 327;
 Best Local Similarity 25.3%; Pred. No. 9.52e+00;
 Matches 22; Conservative 26; Mismatches 35; Indels 4; Gaps 4;
 Db 220 RNDVDMKVALKHGPISTVAIDASHKTSFYSNGVYQKCGNRGOLDHVLVVG-YGE 278
 | : || : : : : || : : : : | : : : : | : : : : | : : : :
 QY 273 REDVQALNISV-PYGPIPVDFQKIRQS-YFASVSYLDTQVGRLLSALDQLANSTIIA 330
 : | : : : : | : : : : | : : : : | : : : : | : : : :
 Db 279 INSEPYWLKPNQGLMGNVIFDVAKN 305
 : | : : : : | : : : : | : : : : | : : : : | : : : :
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356
 RESULT 11
 ID W41993 standard; Protein; 327 AA.
 AC W41993;
 DT 02-JUL-1998 (first entry)
 DE Flea cysteine protease SEQ ID NO:2.
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
 KW Immunoglobulin protease; larvae; host animal.
 OS Siphonaptera.
 PN WO9740058-A1.
 PD 30-OCT-1997.
 PF 24-APR-1997; U06121.
 PR 04-APR-1997; US-042945.
 PR 24-APR-1996; US-639075.
 PR 15-NOV-1996; US-749699.
 PA (HESK-) HESKA CORP.
 PI Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE,
 PI Silver G, Stiegler GL;
 DR WPI: 98-076762/07.
 DR N-PSDB; V04620.
 PT New flea protease genes and proteins - used in vaccine compositions

Query Match	2.6%	Score 107;	DB 1;	Length 4536;
Best Local Similarity	28.0%;	Pred. No. 1.31e+01;		
Matches	23;	Conservative 34;	Mismatches 34;	Indels 5; Gaps 5;
Db	1265	NSLXI-EIPLPFGGKSRDLMLEVTTPALHF-KSVGFHLPSREQVPTFTPKLYQLQ	1322	
	:		:	:
	:		:	:
	:		:	:
	:		:	:
	:		:	:
Qy	187	DVLVDPEGTLDPKQSTEQAIQLLEKMTKSAPSFELAVGYHKHPFRYPK-EFQKLYPLE	245	
	:		:	:
	:		:	:
	:		:	:
	:		:	:
	:		:	:
Db	1323	-VPLLGVLDLSTNYSNLYN-W	1342	
	:		:	:
	:		:	:
	:		:	:
	:		:	:
	:		:	:
Qy	246	NITLAPDFEVPDGLPPVAYNPW	267	
	:		:	:
	:		:	:
	:		:	:
	:		:	:
	:		:	:

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:42:37 2000; MasPar time 8.07 Seconds
Tabular output not generated. 137.498 Million cell updates/sec

Title: >US-09-249-003-3
Description: (1-16) from US09249003.pep
Perfect Score: 117
Sequence: 1 PRELIATSNYPNNIP 16

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.130; Variance 37.308; scale 0.727

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	64	54.7	332	UVR REPAIR PROTEIN.	7.18e-01
2	64	54.7	416	IMPB/MUCB/SAMB FAMILY	7.18e-01
3	64	54.7	416	UV-DAMAGE REPAIR PROTE	7.18e-01
4	63	53.8	192	ESPA.	1.12e+00
5	63	53.8	192	ESPA PROTEIN.	1.12e+00
6	63	53.8	192	SECRETED PROTEIN ESPA.	1.12e+00
7	63	53.8	787	IRON(III) DICITRATE TR	1.12e+00
8	63	53.8	792	IRON(III) DICITRATE TR	1.12e+00
9	62	53.0	434	FLAGELLUM-SPECIFIC ATP	1.73e+00
10	62	53.0	434	ATPASE.	1.73e+00
11	62	53.0	454	MITOCHONDRIAL ELONGATI	1.73e+00
12	62	53.0	471	MITOCHONDRIAL ELONGATI	1.73e+00
13	60	51.3	192	ESPA.	4.10e+00
14	60	51.3	212	MITOCHONDRIAL ELONGATI	4.10e+00
15	60	51.3	259	ORF1.	4.10e+00
16	60	51.3	265	HYPOTHETICAL 31.0 KD P	4.10e+00
17	60	51.3	370	PRE-mRNA SPLICING PROT	4.10e+00
18	60	51.3	755	EXTRACELLULAR MATRIX P	4.10e+00
19	59	50.4	192	ESPA.	6.26e+00
20	59	50.4	1215	DNA FOR SEROTYPE B CAP	6.26e+00

21	58	49.6	160	1	Q9YAF4	160AA LONG HYPOTHETICA	9.51e+00
22	58	49.6	598	3	Q05930	CHROMOSOME XII COSMID	9.51e+00
23	57	48.7	90	10	Q9ZRN9	LECTIN (FRAGMENT).	1.44e+01
24	57	48.7	115	2	Q31841	YOKZ PROTEIN.	1.44e+01
25	57	48.7	256	2	Q34760	YTNP.	1.44e+01
26	57	48.7	331	2	Q9ZCB6	HYPOTHETICAL 36.5 KD P	1.44e+01
27	57	48.7	624	3	Q12246	ORF YOR171C.	1.44e+01
28	57	48.7	1271	2	Q54504	MICROCYSIN SYNTHETASE	1.44e+01
29	57	48.7	2166	14	Q36637	POLYMERASE.	1.44e+01
30	57	48.7	2166	14	Q36635	POLYMERASE.	1.44e+01
31	56	47.9	99	14	Q81975	PAPILLOMAVIRUS (5A3) E	2.17e+01
32	56	47.9	108	1	Q58630	108AA LONG HYPOTHETICA	2.17e+01
33	56	47.9	157	2	Q85405	HYPOTHETICAL 18.0 KD P	2.17e+01
34	56	47.9	186	14	Q9YV77	ORF MSV245 PUTATIVE RN	2.17e+01
35	56	47.9	299	13	Q92153	HAT-2.	2.17e+01
36	56	47.9	401	13	Q93408	HOMEOBOX PROTEIN BIX1.	2.17e+01
37	56	47.9	401	13	Q9YHW3	VACUOLAR-TYPE H-ATPAS	2.17e+01
38	56	47.9	619	5	Q96065	SEMAPHORIN H PRECURSOR	2.17e+01
39	56	47.9	775	11	P70275	SEMAPHORIN H PRECURSOR	2.17e+01
40	56	47.9	775	4	Q15041	COLLAPSIDIN 5 PRECURSOR	2.17e+01
41	56	47.9	785	13	Q42237	CLASS A CALCIUM CHANNE	2.17e+01
42	56	47.9	791	11	Q70368	KIAA1007 PROTEIN (FRAG	2.17e+01
43	56	47.9	1089	4	Q9Y2L0	ENAMELIN PRECURSOR.	2.17e+01
44	56	47.9	1142	6	Q97939	SIMILAR TO HUMAN SREBP	2.17e+01
45	56	47.9	1758	5	Q22830		

ALIGNMENTS

RESULT 1
ID O30670 PRELIMINARY; PRT; 332 AA.
AC O30670;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE UVR REPAIR PROTEIN.
GN UVRX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA PAIK S.H., HANSEN J.N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF014938; AAC63530.1;
DR PFAM; PF00817; IMS; 1.
SQ SEQUENCE 332 AA; 37768 MW; 7F4624B7 CRC32;

Query Match 54.7%; Score 64; DB 2; Length 332;
Best Local Similarity 58.3%; Pred. No. 7.18e-01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 MIDY5QFPRKNI 12
QY 4 LIATSNYPNNIP 15
: | | | : | | |
: | | | : | | |

RESULT 2
ID O64031 PRELIMINARY; PRT; 416 AA.
AC O64031;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE IMPB/MUCB/SAMB FAMILY PROTEIN.
GN YOLE.
OS Bacteriophage SPBC2.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RA LAZAREVIC V., DUESTERHOEFT A., SOLDO B., HILBERT H., MAUEL C.,
RA KARAMATA D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

Dib

1 PRELIANSYPRNNIP 16

RESULT 6
ID O33976 PRELIMINARY; PRT; 192 AA.
AC O33976;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE SECRETED PROTEIN ESPA.
GN ESPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROEC-1;
RX MEDLINE; 97427930.
RA ABE A., KENNY B., STEIN M., FINLAY B.B.;
RT "Characterization of two virulence proteins secreted by rabbit
RT enteropathogenic Escherichia coli, EspA and EspB, whose maximal
RT expression is sensitive to host body temperature.";
RL Infect. Immun. 65:3547-3555(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=413/89-1;
RX MEDLINE; 97045129.
RA EBEL F., DEIBEL C., KRESSE A.U., GUZMAN C., CHAKRABORTY T.;
RT "Temperature- and medium-dependent secretion of proteins by Shiga
RT toxin-producing Escherichia coli.";
RL Infect. Immun. 64:4472-4479(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=413/89-1;
RX MEDLINE; 98389647.
RA KRESSE A.U., SCHULZE K., DEIBEL C., EBEL F., ROHDE M., CHAKRABORTY T.,
RA GUZMAN C.A.;
RT "Pas, a novel protein required for protein secretion and attaching and
RT effacing activities of enterohemorrhagic Escherichia coli.";
RL J. Bacteriol. 180:4370-4379(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RX MEDLINE; 99141579.
RA NOUGAYREDE J.P., MARCHES O., BOURY M., MAINIL J., CHARLIER G.,
RA POHL P., DE RYCKE J., NILON A., OSWALD E.;
RT "The long-term cytoskeletal rearrangement induced by rabbit
RT enteropathogenic Escherichia coli is Esp dependent but intimin
RT independent.";
RL Mol. Microbiol. 31:19-30(1999).
DR EMBL; U80908; AAB71083.1; -
DR EMBL; Y13859; CRA74172.1; -
DR EMBL; AF054421; AAC82338.1; -
SQ SEQUENCE 192 AA; 20534 MW; B3AE5297 CRC32;

Query Match 53.8%; Score 63; DB 2; Length 192;
Best Local Similarity 60.0%; Pred. No. 1.12e+00;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 104 PQEVIDYINPRNDI 118
1 PRELIANSYPRNNI 15

RESULT 7
ID O25487 PRELIMINARY; PRT; 787 AA.
AC O25487;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE IRON(III) DICITRATE TRANSPORT PROTEIN (FECA).
GN HP0807.

OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA LOFTUS K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LECTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AE000592; AAD07854.1; -
DR TIGR; HP0807; -
KW Hypothetical protein.
SQ SEQUENCE 787 AA; 88946 MW; BB69068C CRC32;

Query Match 53.8%; Score 63; DB 2; Length 787;
Best Local Similarity 53.3%; Pred. No. 1.12e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 536 KELLFFNYORSYIP 550
2 RELIANSYPRNNIP 16

RESULT 8
ID O92L37 PRELIMINARY; PRT; 792 AA.
AC O92L37;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE IRON(III) DICITRATE TRANSPORT PROTEIN.
GN FECA_2.

OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99120557.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001505; AAD06315.1; -
SQ SEQUENCE 792 AA; 89711 MW; C7507904 CRC32;

Query Match 53.8%; Score 63; DB 2; Length 792;
Best Local Similarity 53.3%; Pred. No. 1.12e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 536 KELLFFNYORSYIP 550
2 RELIANSYPRNNIP 16

RESULT 9
ID O92JJ3 PRELIMINARY; PRT; 434 AA.
AC O92JJ3;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

```

DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE FLAGELLUM-SPECIFIC ATP SYNTHASE.
GN FLII.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=J99;
RX MEDLINE; 99120557.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA TOMLINNO P.J., CARUSO A., ORIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RL Nature 397:176-180(1999).
DR EMBL; AE001554; AAD06888.1; -. A4523E49 CRC32;
SQ SEQUENCE 434 AA; 47703 MW; 47703 MW; 221EA346 CRC32;

Query Match 53.0%; Score 62; DB 2; Length 434;
Best Local Similarity 53.3%; Pred. No. 1.73e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 334 SRELTGYIYPPINI 348
QY 1 PRELIAYSNYPNNI 15
:||||| :|||
:||||| :|||

RESULT 10
ID 054374 PRELIMINARY; PRT; 434 AA.
AC 054374;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE ATPASE.
GN FLII.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CCUG 17874;
RA PORWOLLIK S., O'TOOLE P.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75584; AAB93498.1; -.
DR PFAM; PF00006; ATP-synt_ab; 1.
SQ SEQUENCE 434 AA; 47592 MW; 221EA346 CRC32;

Query Match 53.0%; Score 62; DB 2; Length 434;
Best Local Similarity 53.3%; Pred. No. 1.73e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 334 SRELTGYIYPPINI 348
QY 1 PRELIAYSNYPNNI 15
:||||| :|||
:||||| :|||

RESULT 11
ID 092T91 PRELIMINARY; PRT; 454 AA.
AC 092T91;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU.
GN TUF4 OR T419.19 OR T5J8.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA PARNEILL L.D., GNOJ L., DE LA BASTIDE M., HAMEED A., HABERMANN K.,
RA SCHUTZ K., HUANG E., GOTTESMAN T., DEDHIA N.N., MCCOMBIE W.R.;
RT "Genomic sequence of BAC T419 from Arabidopsis thaliana, Chromosome
RT IV, near 16.6 cM.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RA DE LA BASTIDE M., GNOJ L., HABERMANN K., HUANG E.N., GOTTESMAN T.,
RA KAPLAN N., LODHI M., JENSEN K., HAMEED A., SCHUTZ K., MARTIENSSEN R.,
RA DEDHIA N., PARNEILL L.D., MCCOMBIE W.R.;
RT "Arabidopsis thaliana BAC T5J8 from chromosome IV, short arm.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069442; AAC79113.1; -.
DR EMBL; AC004044; AAD15337.1; -.
DR HSSP; P02990; IEFU.
DR MENDEL; 34382; Arath; TufA; 34382.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding.
SQ SEQUENCE 454 AA; 49410 MW; 3B0B22FE CRC32;

Query Match 53.0%; Score 62; DB 10; Length 454;
Best Local Similarity 40.0%; Pred. No. 1.73e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 210 RELISFYKPGDDIP 224
QY 2 RELIAYSNYPNNIP 16
:||||| :|||
:||||| :|||

RESULT 12
ID 039206 PRELIMINARY; PRT; 471 AA.
AC 039206;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU PRECURSOR.
GN TUF4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RX MEDLINE; 96145515.
RA KUHLMAN P., PALMER J.D.;
RT "Isolation, expression, and evolution of the gene encoding
RT mitochondrial elongation factor Tu in Arabidopsis thaliana.";
RL Plant Mol. Biol. 29:1057-1070(1995).
[2]
RN SEQUENCE OF 359-471 FROM N.A.
RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;
RA COOKE R., LAUDIE M., RAYNAL M., DELSENY M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89227; CAA61511.1; -.
DR EMBL; F14375; CAA23078.1; -.
DR HSSP; P02990; IEFU.
DR MENDEL; 13559; Arath; TufA; 13559.
DR PROSITE; PS00301; EFATOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
KW Transit peptide; Elongation factor; Protein biosynthesis;
KW GTP-binding.
FT TRANSIT 1 72 POTENTIAL.
FT CHAIN 73 471 MITOCHONDRIAL ELONGATION FACTOR TU.
SQ SEQUENCE 471 AA; 51384 MW; 849FCE59 CRC32;

Query Match 53.0%; Score 62; DB 10; Length 471;
Best Local Similarity 40.0%; Pred. No. 1.73e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 227 RELSFKYKPGDDIP 241
   |||::: ::||
QY 2 RELIAYSNYPRNNIP 16

RESULT 13
ID Q47184 PRELIMINARY; PRT; 192 AA.
AC Q47184;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE ESPA.
DE ESPA.
GN ESPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ENTEROPATHOGENIC SEROTYPE O127:H6;
RX MEDLINE; 96310369.
RA KENNY B., LAI L., FINLAY B., DONNENBERG M.S.;
RT "EspA, a protein secreted by enteropathogenic Escherichia coli, is
RT required to induce signals in epithelial cells.";
RL Mol. Microbiol. 20:313-323(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA ELLIOTT S.J., WAINWRIGHT L.A., MCDANIEL T.K., JARVIS K.G., DENG Y.K.,
RA LAI L.C., MCNAMARA B.P., DONNENBERG M.S., KAPER J.B.;
RL Mol. Microbiol. 0:0-0(1998).
RN [3]
RP SEQUENCE OF 1-191 FROM N.A.
RC STRAIN=30;
RA FRANKEL G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=30;
RX MEDLINE; 99068022.
RA NEVES B.N., KNUSTON S., TRABULSI L.R., SPERANDIO V., KAPER J.B.,
RA DOUGAN G., FRANKEL G.;
RT "Molecular and ultrastructural characterisation of EspA from different
RT enteropathogenic Escherichia coli serotypes.";
RL FEMS Microbiol. Lett. 169:73-80(1998).
DR EMBL; Z54352; CAA91163.1; -
DR EMBL; AF022236; AAC38394.1; -
DR EMBL; AJ225019; CAA12349.1; -
SQ SEQUENCE 192 AA; 20469 MW; 8D6D33B9 CRC32;

Query Match 51.3%; Score 60; DB 2; Length 192;
Best Local Similarity 60.0%; Pred. No. 4.10e+00;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 104 PDEVISYINDPRNDI 118
   ||:|:| |||:|
QY 1 PRELIAYSNYPRNNI 15

RESULT 14
ID Q9ZR06 PRELIMINARY; PRT; 212 AA.
AC Q9ZR06;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU (FRAGMENT).
GN TUFA.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Gentianales; Apocynaceae;
OC Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.

RA SCHROEDER G., LURZ G., SCHROEDER J.;
RT "Mitochondrial elongation factor Tu from Catharanthus roseus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ130966; CAA10267.1; -
DR HSSP; P02990; LEFU.
DR MENDEL; 34753; Catro:Tufa;34753.
KW Elongation factor.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23787 MW; 69A1E059 CRC32;

Query Match 51.3%; Score 60; DB 10; Length 212;
Best Local Similarity 40.0%; Pred. No. 4.10e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 11 RELSFKYKPGDEIP 25
   |||::: ::||
QY 2 RELIAYSNYPRNNIP 16

RESULT 15
ID Q44477 PRELIMINARY; PRT; 259 AA.
AC Q44477;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ORF1.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC Azotobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CA;
RA MYLONA P.V., RICKE S.C., BISHOP P.E.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39675; AAA82617.1; -
DR PFAM; PF00005; ABC_tran; 1.
SQ SEQUENCE 259 AA; 28537 MW; C337B05C CRC32;

Query Match 51.3%; Score 60; DB 2; Length 259;
Best Local Similarity 41.7%; Pred. No. 4.10e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 96 RELVGFGRYPWH 107
   |||::: ||:|
QY 2 RELIAYSNYPRN 13

Search completed: Wed Apr 19 21:44:12 2000
Job time : 95 secs.

```

WAP5024 (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Oct 4 13:48:44 1995; MasPar time 1618.89 Seconds
1044.595 Million cell updates/sec

Tabular output not generated.

Title: >US-08-345-212-1
Description: (1:2297) from US08345212.seq
Perfect Score: 2297
N.A. Sequence: 1 CGGCTGTGTGGAGTCTT.....TAATCATGTTTCTTTTTC 2297
Comp: GCCGACACACCGCTCAGAA.....ATTAGGTACAAGAAAAAGG

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 493065 seqs, 368106350 bases x 2

Database: emb1-new6

- 1 BCT
- 2 EST1
- 3 EST2
- 4 EST3
- 5 EST4
- 6 EST5
- 7 FUN
- 8 INV1
- 9 INV2
- 10 MAM
- 11 ORG
- 12 PLN
- 13 PRI
- 14 PRO
- 15 ROD
- 16 STS
- 17 SYN
- 18 UNC
- 19 VRT
- 20 VIR

Database: EST
21 EST1
22 EST2
23 EST3
24 EST4
25 EST5

Database:

genbank89
79 BCT1
80 BCT2
81 BCT3
82 BCT4
83 BCT5
84 BCT6
85 INV1
86 INV2
87 INV3

- 26 EST6
- 27 EST7
- 28 EST8
- 29 EST9
- 30 EST10
- 31 EST11
- 32 EST12
- 33 EST13
- 34 EST14
- 35 EST15
- 36 EST16
- 37 EST17
- 38 EST18
- 39 EST19
- 40 EST20
- 41 EST21
- 42 EST22
- 43 EST23
- 44 EST24
- 45 EST25
- 46 EST26
- 47 EST27
- 48 EST28
- 49 EST29
- 50 EST30
- 51 EST31
- 52 EST32
- 53 EST33
- 54 EST34
- 55 EST35
- 56 EST36
- 57 EST37
- 58 EST38
- 59 EST39
- 60 EST40
- 61 EST41
- 62 EST42
- 63 EST43
- 64 EST44
- 65 EST45
- 66 EST46
- 67 EST47
- 68 EST48
- 69 EST49
- 70 EST50
- 71 EST51
- 72 EST52
- 73 EST53
- 74 EST54
- 75 EST55
- 76 EST56
- 77 EST57
- 78 EST58

88 INV4
89 INV5
90 MAM1
91 MAM2
92 PAT1
93 PAT2
94 PHG
95 PLN1
96 PLN2
97 PLN3
98 PLN4
99 PLN5
100 PLN6
101 PLN7
102 PRI1
103 PRI2
104 PRI3
105 PRI4
106 PRI5
107 PRI6
108 PRI7
109 PRI8
110 PRI9
111 ROD1
112 ROD2
113 ROD3
114 ROD4
115 ROD5
116 ROD6
117 ROD7
118 STR
119 STS1
120 STS2
121 STS3
122 STS4
123 SYN
124 UNA
125 VRL1
126 VRL2
127 VRL3
128 VRL4
129 VRL5
130 VRL6
131 VRT1
132 VRT2
133 VRT3

Database:

genbank-new6

134 BCT
135 EST1
136 EST2
137 EST3
138 EST4
139 EST5
140 EST6
141 EST7
142 EST8
143 INV
144 MAM
145 PHG
146 PLN
147 PRI
148 ROD
149 STS

150 STR
151 SYN
152 UNA
153 VRL
154 VRT

Database: u-emb143_89
155 ALL

Statistics: Mean 12.082; Variance 3.168; scale 3.813

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2297	100.0	2297	107	HUMIDSX	Human iduronate 2-sul	0.00e+00
2	1096	47.7	1750	113	MUSIDS	Mus musculus iduronat	0.00e+00
3	996	43.4	1120	107	HUMIDSGEN9	Homo sapiens iduronat	0.00e+00
4	995	43.3	36845	107	HUMIDS	Homo sapiens iduronat	0.00e+00
5	469	20.4	615	142	H06357	yl79h02.r1 Homo sapie	0.00e+00
6	469	20.4	615	3	HS357151	yl79h02.r1 Homo sapie	0.00e+00
7	434	18.9	1831	117	S64545	iduronate sulfatase I	0.00e+00
8	412	17.9	469	2	HS179158	yl85c04.r1 Homo sapie	0.00e+00
9	412	17.9	469	142	H05179	yl85c04.r1 Homo sapie	0.00e+00
c 10	369	16.1	378	56	T08577	EST06469 Homo sapiens	0.00e+00
c 11	360	15.7	484	38	R15374	yl90e05.r1 Homo sapie	0.00e+00
c 12	332	14.5	439	142	H06300	yl79h02.s1 Homo sapie	0.00e+00
c 13	332	14.5	439	3	HS300154	yl79h02.s1 Homo sapie	0.00e+00
c 14	321	14.0	327	26	HSC0SC011	H. sapiens partial cD	0.00e+00
c 15	319	13.9	423	71	T71926	ye07a10.r1 Homo sapie	0.00e+00
c 16	291	12.7	363	75	T87633	ye07a10.s1 Homo sapie	0.00e+00
c 17	290	12.6	297	31	HSC2PG011	ye07a10.s1 Homo sapie	0.00e+00
c 18	287	12.5	289	30	HSC2VA071	H. sapiens partial cD	0.00e+00
c 19	275	12.0	284	30	HSC2UC052	H. sapiens partial cD	0.00e+00
c 20	274	11.9	319	52	R64400	yl15h10.r1 Homo sapie	0.00e+00
c 21	258	11.2	261	30	HSC2UC051	H. sapiens partial cD	0.00e+00
c 22	258	11.2	310	46	R42115	yl90e05.s1 Homo sapie	0.00e+00
c 23	251	10.9	251	26	HSC0GA021	H. sapiens partial cD	0.00e+00
c 24	250	10.9	250	26	HSC0GA011	H. sapiens partial cD	0.00e+00
c 25	233	10.1	288	52	R64354	yl15a10.r1 Homo sapie	7.38e-304
c 26	231	10.1	233	24	HSAAABLT	H. sapiens putatively	1.03e-300
c 27	229	10.0	536	107	HUMIDSGEN1	Homo sapiens iduronat	1.42e-297
c 28	229	10.0	279	5	HS695141	yl85c09.r1 Homo sapie	1.42e-297
c 29	229	10.0	279	139	R79695	yl85c09.r1 Homo sapie	1.42e-297
c 30	221	9.6	237	30	HSC2VA072	H. sapiens partial cD	5.09e-285
c 31	218	9.5	238	26	HSC0SC012	H. sapiens partial cD	2.55e-280
c 32	214	9.3	282	52	R64401	yl15h10.s1 Homo sapie	4.67e-274
c 33	214	9.3	344	52	R64355	yl15a10.s1 Homo sapie	4.67e-274
c 34	205	8.9	322	108	HUMISS05	Homo sapiens iduronat	5.41e-260
c 35	205	8.9	490	107	HUMIDSGEN5	Homo sapiens iduronat	5.41e-260
c 36	186	8.1	202	65	T49494	ya75h11.s1 Homo sapie	1.98e-230
c 37	184	8.0	196	26	HSC0GA022	H. sapiens partial cD	2.49e-227
c 38	182	7.9	388	107	HUMIDSGEN3	Homo sapiens iduronat	3.13e-224
c 39	182	7.9	750	147	HUMIDS02	Homo sapiens iduronat	3.13e-224
c 40	182	7.9	750	13	HSIDS02	Homo sapiens iduronat	3.13e-224
c 41	182	7.9	298	108	HUMISS03	Homo sapiens iduronat	3.13e-224
c 42	178	7.7	290	108	HUMISS06	Homo sapiens iduronat	4.84e-218
c 43	178	7.7	379	107	HUMIDSGEN6	Homo sapiens iduronat	4.84e-218
c 44	176	7.7	404	107	HUMIDSGEN8	Homo sapiens iduronat	5.97e-215


```
QY 961 AACAATCAGTGTGCCGATGGTCGAATTCCTGTGGACTTTCACGGGAAATCGCCAGAG 1020
Db 1021 ctacttgcctctgtgcatatttgatgatacaacaggtcgccgcctcttgagtgcttga 1080
QY 1021 CTACTTTGCTCTGTGTATATTGGATACACAGGTGCGCGCCTCTTGAGTGCCTTGA 1080
Db 1081 cgatcttcagctggccacagcaccatcattgcatctacctcggatcatggtggctct 1140
QY 1081 CGATCTTCAGCTGGCGACAGCACCATCAITTCATTCCTCGGATCATGGTGGCTCT 1140
Db 1141 aggtgaactggagatggccaaatcacagcaatttggatgttgcctaccatgtccct 1200
QY 1141 AGGTGAACATGGAGATGGGCCAAATACAGCAATTTTGATGTGTCTGCTACCGATGTTCCCT 1200
Db 1201 gatctctatgttctcgaagacgcttcacttcggagcaggcgagaagctttccc 1260
QY 1201 GATATTTCTATGTCTCGAAGACGGCTTCAC TTCGGAGCGAGCGAAGCTTTTCCC 1260
Db 1261 ttacctcgaccttttgatccgcctcacagttgatggagccagcaggcaatccatgga 1320
QY 1261 TTACCTCGACCCCTTTTGATTCGCCCTCACAGTTGATGGAGCGCAGCGCAATCCATGGA 1320
Db 1321 ccttgtggaactgtgtctcttttcccacgtggtggaactgcaggatgcaggttcc 1380
QY 1321 CTTTGTGGAAC TTGTCTCTTTTCCACGCTGGCTGGACTTCGAGGACTGCAGGTTCC 1380
Db 1381 acctgcgtgcccgcttccttcttccatttcacgttgagctgtgcagagaaggcaaaccttc 1440
QY 1381 ACCTGCGTGCCTGCTTCCTTCATTTTCAGCTGTGAGCTGTGCAGAGAAGCAAGACCTTCT 1440
Db 1441 gaagcatttccagctcgtgacttgaagaggatcgtacctccctcgttaatcccgatga 1500
QY 1441 GAAGCATTTTCGATTCCTGCTGACTTGAAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA 1500
Db 1501 actgatgcctatgacagatgaccccgcttcagacatccctcagtggaatctgacaa 1560
QY 1501 ACTGATTCCTATAGCCAGTATGCCGCTTCAGACATCCCTCAGATCCCTCAGTGAATTCGCACAA 1560
Db 1561 gccagatttaaaagataaagatcatggctattccatcacaccatagactatagtta 1620
QY 1561 GCCAGTTTAAAGATATAAAGATCATGGCTATTCATACGCCACCATAGACTATAGTA 1620
Db 1621 tactgtgtgggttgcttcaatcctgatgaattctagctaactttctgcacatccatgc 1680
QY 1621 TACTGTGTGGGTGCTTCATTCCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680
Db 1681 aggggaactgtatgttgattctgaccttgaccatgcaggatcacatatgtataatgattc 1740
QY 1681 AGGGGAAC TGATTTTGTGCAATTCACCCATTCGAGGATCAAAATATGTATAATCATTC 1740
Db 1741 ccaaggtggagatcttttcagcttgtgatgccttgagttttcccaaccatggatggcaa 1800
QY 1741 CCAAGGTGGAGATCTTTTCCAGTTGCTGATCCCTTCAGTTCCTGCTTTTCCCAACCATGGCNA 1800
Db 1801 atgtgatgtgctccctccagctggtgagaggaggttagagctggtcgttttgtatt 1860
QY 1801 ATGTGATGTCTCCCTCCAGCTGGTGAGAGGAGGAGTTAGAGCTGCTGCTTTTGTGATT 1860
Db 1861 acccataattggaagcagcctgagggctagttaatccaaacatgcatacaaatgttg 1920
QY 1861 ACCCATATATTTGGAGACGCTGAGGCTGAGGCTAGTTAATCCAAACATGCATCAACAAATTTGG 1920
Db 1921 cctgagaatgatgaacagcaaaccttttcgtttagctttattataaaattataatggt 1980
```

```
QY 1921 CTGGAATATGTACAGCCAAACCTTTTGTAGTCTTTATTAATAATTATAATTGGT 1980
Db 1981 aattggaccagtttttttttaatttcctctttttaaacaagttacgcttatttactg 2040
QY 1981 AATTGGACCAGTTTTTAAATTTCCCTCTTTTAAACAGCTTACGGCTTATTTACTG 2040
Db 2041 ataataacaagcaacaactcaagttatgtcatccttttgatcacgaagaccataca 2100
QY 2041 AATAATACAAAGCAACAACTCAAGTTATGTACCTTTGGATACGAAGACCATACA 2100
Db 2101 taatacaaaacataacattatcacacaagaatacctttcattattgttgggaatttagtc 2160
QY 2101 TAATACCAACATACATTAATACAAAGAAATCTTCATTATTGTGGAATTTAGTGC 2160
Db 2161 atttcaaaaagaatcatatatacaaaactaggccaccacactaaagtcctcgatttttgt 2220
QY 2161 ATTCAAAAAGTAATCATATATCAAACTAGGCACCACTAAGTTCCTGATTATTTGTT 2220
Db 2221 tataatttaataatatcttatgagccctatatattcaaaatattattgtaaactgtaa 2280
QY 2221 TATAATTTAATAATATCTTATGAGCCCTATATATCAAAATATTATGTTAAACAICTAA 2280
Db 2281 tcatgtttctttttcc 2297
QY 2281 TCATGTTTCTTTTCC 2297
RESULT 2
LOCUS MUSIDS 1750 bp mRNA ROD 19-AUG-1993
DEFINITION Mus musculus iduronate sulfatase (Ids) mRNA, complete cds.
ACCESSION L07921
KEYWORDS iduronate sulfatase.
SOURCE Mus musculus (library: lambda ZAP II of Stragene) female adult
thymus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE 1 (bases 1 to 1750)
AUTHORS Daniele,A., Faust,C.J., Herman,G.E., Natale,P.D. and Ballabio,A.
TITLE Cloning and characterization of the cDNA for the murine iduronate
sulfatase gene
JOURNAL Genomics 16, 755-757 (1993)
MEDLINE 93315172
COMMENT NCBI gi: 349713
FEATURES location/Qualifiers
source 1..1750
/organism="Mus musculus"
/dev_stage="adult"
/sequenced_mol="cDNA to mRNA"
/sex="female"
/tissue_type="thymus"
/tissue_lib="lambda ZAP II of Stragene"
21..1712
/gene="Ids"
/EC_number="3.1.6.13"
/note="NCBI gi: 349714"
/codon_start=1
/product="Iduronate sulfatase"
CDS
/translation="MVVDVTLPGNHVQSHRTCARIMRQLSFLILGFSFIALESAAQ
GNSATDALNILLIVDDLRPSLCYCKIVRSPNIDOLASHSVLFQNAFAQQAVCAPS
RVSLTGRPDTRLIDFNFSYHRVHSGNSTIPQIFKENGIVTMVGKVFHPGIGSNH
SDPYPSWSFYPHPSPFFLAVGYHKPFPKPKFQKLYPLENI TLAPDPHPDPLPP
EATRLIEKMTSGSPFFLAVGYHKPFPKPKFQKLYPLENI TLAPDPHPDPLPP
VAYNPWMDIREDDVQALNISVPYCP IPEDFQFKIRQSYFASVSYLDITQVGHVLSALD
```


DLR LAHNT I I A F T S D H G W A L G E H G E W A K Y S N E D V A T R V P L M L Y V P G R T A P L P A Q G K L I
F P Y R D P F P A S D M W D A C R H T L V E L V S L F P T L A G I A G L P V L I G A P S L L F M L S F A E K A
R I F R S I C S S W T K R S Q T C L V P R E L T A Y S O Y P R A D F P Q W N S K P T L N D I T R S W D I Y T H V
D Y R Y T V W V G F D P S E F L A N F S D I H A G E L Y F V D S D P L Q D H N V Y N D S Q H G S L L L H S L R P "

BASE COUNT	400 a	481 c	414 g	455 t	ORIGIN				
DB 113;	Score	1096;	Match	84.4%;	QryMatch	47.7%;	Pred. No.	0.00e+00;	
Matches	1382;	Conservative	0;	Mismatches	250;	Indels	6;	Gaps	5;
Db	109	tgttctaggtcctcttgcatacgccctggagtcgcgagcggaagcaagcaactctgctacag	168						
Qy	168	TGTTTTCAGCTCCGCTTGGCTCGCCTCGGATCCGAAACCGACGCCAACTCGACACAG	227						
Db	169	atgctctgaacatcctctgatacttggatgatactcccgccctccctggcgtgttatg	228						
Qy	228	ATGCTCTGAACGTTCTTCTCATCTGCTGGATCATCGCGCCCTCCCTGGCGTGTATG	287						
Db	229	gagacaagctagtgaggtccccgaacattgaccagctggcatcccatagagctctcttc	288						
Qy	288	GGGATAACCTGTGTAGGTCCCAATAATTGACCACTGGCATCCACAGCCTCCTCTTCC	347						
Db	289	agaatgcttttgtcagcaagcagtgctgcaccaagtgcagtgctctctccacggga	348						
Qy	348	AGATGCTTTGGGAGCAGACAGATGTGGCGCCGAGCGCGTGTCTTCTCATCGCA	407						
Db	349	ggagcctgacaccaccgcctgtatgactcaattctactgagaggtacattctggaa	408						
Qy	408	GAGAGCTTGACACACCGCGCTGTACGACTTCAACTCCTACTGGAAGGTGCACCTG	467						
Db	409	attttccaccatcccccaacttcaaggagaatggctacgtgacatgctcggtggga	468						
Qy	468	ACTTCTCCACCATCCCCAGTACTTCAAGGAGATGCGTATGTGACCATGTCTCGTGG	527						
Db	469	aagctcttaccctggtatatacttccaaatcacagtgatgatatacctatagctggtctt	528						
Qy	528	AAGCTTTTACCCCTGGGATATCTTCAACCATACCGATGATTCCTCGTATAGTGTGCTT	587						
Db	529	tteccattatcatccatcctcagagaaatgaaacacataagcgtgaaggacaa	588						
Qy	588	TTCCACCTTATCATCTTCTCTGAGAGTATGAAACACTAAGACATGTCGAGGGCCAG	647						
Db	589	atggaaagctccacccaatcgtttgtcctgctgagtggtgcagatgctcgaagta	648						
Qy	648	ATGGAGAACTTCATGCCAACTGCTTCCCTCTGGATGTGCTGCATGTTCCCGAGGGCA	707						
Db	649	ccttgctgacaaacagacagacagaagccatcagttgttgaaagatgaagacat	708						
Qy	708	CTTCTCTGCAAAA CAGAGCACTGAGAACGCATACAGTTGTTGGAAAGATGAAACGT	767						
Db	709	caggtagctcttctcctgagttgggtaccacaagcgcgatatcccttcagatacc	768						
Qy	768	CAGCCAGTCTCTTCTCTCGCCGTTGGGTATTAAGCCACACATCCCTTCAGATACC	827						
Db	769	caaaggaatttcagaagttgatcccttgaaacataaaccctgctcctgatccccatg	828						
Qy	828	CCAAGGAATTCAGAAGTTGATCTCTTGGGAAACATCAACCTTGGCGCCGATCCCGAGG	887						
Db	829	tctcgtatgctcaccactgtagctacaaacccctggatggatatacaggagagggag	888						
Qy	888	TCCCTGATGGCCTACCCCTGTGGCTACAAACCCCTGGATGCATCAGGCAACGGGAAG	947						
Db	889	atgtccaaagcttaaacatcagttgctcctatggaccaattcctgaggtatttcacggga	948						

Qy 948 ACGTCCACAGCCTTAAACATACAGTGTGGCGATGATGCTCCATATCTGTGGAGCTTTTCAGCGGA 1007

Db 949 aaatccgcagagctaactttgcttctgtgctctatttggatactcaagtcggaacatgtttt 1008

Qy 1008 AAATCCGCCAGAGTACTTTGCCCTCTGTCTCATATTTGGATACACAGTGCGGCGCTCT 1067

Db 1009 tgaagtctttggatgatctctgctctagccacacacaattattgtttcacatctgatc 1068

Qy 1068 TGAGTCTTTTGGAGATCTTCAGCTGGCCACACAGCACCATCAITTCATTTACCTCGGATC 1127

Db 1069 acggtatgggacctaggtgaacacgaggaatgggccaagtacagcaactttgatgttgeta 1128

Qy 1128 ATGGGTGGGCTTAGCTGAACTGGAGAAATGGGCCAATAACAGCAATTTTGATGTTGCTA 1187

Db 1129 ccgctgtgccctgatgctctatgtccctgaaaggacggcccaacttctcggcgagcc 1188

Qy 1188 CCCATGTCTCCCTGATATTCATGTTCTCTGGAAGGACGGCTTCAC TTCGGAGGCAAGG 1247

Db 1189 agaagctcttccataccggaagacccttggatccggcctcggatggatggatgcaggca 1248

Qy 1248 AGAAGCTTTTCCCTTACCTACGACCCCTTTTGATTTCGCGCTCACAGTTGATGGAGCAGGCA 1307

Db 1249 ggcacacgagagacctcgtagaacttgtatctcttcccacgctcgtcgtgacttgcag 1308

Qy 1308 GGCATTCATCGAACCTTTGGAACTTTGTGTCTCTCTTTTCCACGCTGGCTGGACTTGCAG 1367

Db 1309 gactgccagt--cctcctcgtggcccacccctcttttctcatgttgacttttgacagaag 1367

Qy 1368 GACTGCGAGTTTCCACCTCGCTGGCCGCTTCCTATTTCAGCTTGAGCTTGCAGAGAG 1427

Db 1368 gccagaattctcagaagcatttgacgtccatgacttggagaagagcagaattgttg 1427

Qy 1428 GCAAGAACTTCTGAAGCAATTTTCGATTCCGTGACTTGGAAAGGATCCGTACCTCCCTG 1487

Db 1428 gta----ccccggagttgatcgctctatagccagtaccceggcctcgagattccctcagt 1485

Qy 1488 GTAATCCCGTGAACTGATTCGCTATAGCCAGTATCCCGGGCTTCAGACATCCCTCAGT 1547

Db 1486 ggaattctgacacgccaagctaaatgatataag--lcatgggatat--ctacagcacg-- 1542

Qy 1548 GCAATTTCTGCAAGCCGAGTTTAAAGATATAAGATCATGGGCTATTCCATACGCCACA 1607

Db 1543 tagactacaggtacacagtgtggttgcttcgactcagtcagtgaattcttgctaacattt 1602

Qy 1608 TAGACTATAGTATACTGTGTGGGTTGGCTTCAATCCTCATGAATTTCTAGCTAACTTTT 1667

Db 1603 ctgataccatgcggggaactctatttgcattctgaccactcaggtacgagatcacaatg 1662

Qy 1668 CTGAGATCATGACGGGAGACTGTATTTTGTGANTTGCACCCATTCGAGAGTCAATA 1727

Db 1663 tatataatgactcccacaatgaggtctcttcccactcactgaggtctgagttctgccaag 1722

Qy 1728 TGATATAATGATTCCCAAGTGGAGATCTTTTCCAGTTGTTGTATGSCCTTGAGTTTGCACA 1787

Db 1723 cctcagtgggcacaatatg 1740

Qy 1788 CCATGGATGGCAATCTG 1805

RESULT	3
LOCUS	HUMIDSCN9 1120 bp DNA PRI 18-MAR-1994
DEFINITION	Homo sapiens iduronate-2-sulfatase (IDS) gene, complete cds.
ACCESSION	L13329
KEYWORDS	IDS gene; iduronate 2-sulfatase.

```
SEGMENT 9 of 9
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
REFERENCE 1 (bases 1 to 1120)
AUTHORS Wilson,P.J., Meaney,C.A., Hopwood,J.J. and Morris,C.P.
TITLE Sequence of the human iduronate 2-sulfatase (IDS) gene
JOURNAL Genomics 17 (3), 773-775 (1993)
MEDLINE 94063929
COMMENT NCBI gi: 405203
FEATURES
    Location/Qualifiers
        1..1120
            /organism="Homo sapiens"
            /sequenced_mol="DNA"
            /tissue_lib="lambda Charon 35, ATCC No. LAOXN101"
            join(L13321:332..434,L13322:157..293,L13323:82..259,
            L13324:131..219,L13325:125..325,L13326:88..258,
            L13327:99..225,L13328:129..302,L13329:128..600)
            /gene="IDS"
            /map="Xq27.3-q28"
            /note="NCBI gi: 405205"
            /codon_start=1
            /product="iduronate-2-sulfatase"
            /translation="MPPRTGRGLIWLGIWLVSSVCVALGSETQANSTTDLNVLIV
            DDLRPLSCYGGKLVSPNIDQLASHLLFQNAFAQAVCAVRSEFLTGRDPTRL
            YDFNSYMRVHAGNFSTIPQYFKENGYYTMSVGVKHFPISSNHTDDSPYSWGFPPYHP
            SSEKYMTKTCRGDGLHIANLCFVDVLDVPEGTLPDKOSTEQDIOLEKMKTSASP
            FFLAVGYHKPHIPRYPEKFKLYLENITLAPDFEYDGLPPVAYNPWDINQEDV
            QALN1SVYGP1PVDVFORK1ROSYFASVYLDTVQVRLSALDDIQLANSTIIAFTSD
            HGMALGHEGEMAKTSNEFVATVPLIFYPVGRATSLPEAGEKILFYLDPEFDSQIME
            PGRQMDVLVLSLEPTLAGLAGLVPPCPVPSFHWELCREKNLLKHFREFDLEED
            PLYPNRELIAISQYPRSDIPQWNSDKPSIKDKIKMGYSIRTIIDYRYTVWVGFNP
            EFLANFSDIHAGELYFVDSPLQDHNMYNDSSQGDLEQLLMP"
            order(L13328:303..404,1..127)
            /gene="IDS"
            /map="Xq27.3-q28"
            /number=8
            128..>1120
            /gene="IDS"
            /map="Xq27.3-q28"
            /number=9
            /evidence=experimental
BASE COUNT 318 a 232 c 212 g 358 t
ORIGIN

Db 125 caggcaggcaatccatgacgtgtggaactgtgtctctttttccacgctgagctgac 184
      |||
Qy 1302 CAGGCGAGCAATCCATGACCTGTGGAACTGTGTCTCTTTCCACCGCTGGCTGCA 1361

Db 185 ttgcaggactgcaggttccactcgtcgcctccttccatttccattgagctgtgca 244
      |||
Qy 1362 TTGCAGGACTGCAGGTTCCACCTCGCTCCGCCCTTCCTTCATTTCAGCTTGAGCTGCA 1421

Db 245 ggaagcgaagaacctctgaagcatttccgattccgctgacttgggaagagagctacc 304
      |||
Qy 1422 GAGAAGCAAGAACCTCTCGAACATTTTCGATTTCGCTGACTTGGGAAGGATCCGTACC 1481

Db 305 tcccttgtaatccccgtaagctgattgctatagccagatcccccgcttcagacatcc 364
      |||
Qy 1482 TCCCTGTATCCCCGTGACTGATTGCTTATAGCAGTATCCCGCGCTTCAGACATCC 1541
```

```
Db 365 ctcaagtgaattctgacaaagccaggtttaaaagatataaagatcatgggtgattccatac 424
      |||
Qy 1542 CTCAGTGGATCTTGACAGCCGAGTTTAAAGAGATATAAGATCATGGGCTATTTCATAC 1601

Db 425 gcaccatagactataggtataactgtgtgggttggtctcaatcctcgtgatgaattcttagta 484
      |||
Qy 1602 GCACCATAGACTATAGGTATAGTGTGTGGGTGGCTCAATCCCTGATGAATTTCTAGTCA 1661

Db 485 acttttctgacatccatgcaggggaactgtatttctggtgattctgacccattgcagatc 544
      |||
Qy 1662 ACTTTTCTGACATCCATGCAGGGGAATCTATTTTGTGGATTCTGACCATTCGAGGATC 1721

Db 545 acaatgtataatgattcccaagtgagatctttccagttgttgatgccttgagttt 604
      |||
Qy 1722 ACAATATGTATAATGATTTCCCAAGGTGGAGATCTTTCCAGTGTGTGATGCTTGAGTTT 1781

Db 605 tgcacaacctgagtgcaaatgtgatgtcctccctccagctggtgagaggaggttag 664
      |||
Qy 1782 TGCACAACCTGAGATGCAAAATGTGATGTCTCCCTCCAGCTGGTGAGAGAGGAGTTAG 1841

Db 665 agctggtcgtttgtgattaccataatattggaagcagcctgaggtcagttaatccaa 724
      |||
Qy 1842 AGCTGCTGCTTTTGTGATTACCCATAATATTGGAAGCAGCTCGAGGCTAGTTAATCCAA 1901

Db 725 acatgcatacaaatgtgctgagatatagtacaagccaaaccttttcgttagcttt 784
      |||
Qy 1902 ACATGCATCAACAATTTGGCTCAGAAATGTGATGTCTCCCTCCAGCTGGTGAGAGAGGAGTTAG 1961

Db 785 attaaattataaattgtaattgacaggttttttttaattccctctttttaaacc 844
      |||
Qy 1962 ATTAAATTTAATATGTAATTTGGACAGCTTTTATTTTAAATTCCTCTTTTAAAC 2021

Db 845 agtcagcgttattactgaataatacaaaagcaaaactcaagttatgcatacctt 904
      |||
Qy 2022 AGTTAGCGCTATTACTGAAATAATAACAAAGCAAACTCAAGTTATGTCATACCTT 2081

Db 905 tggatagaagaccatacatataaaccatacaataacattatatacaaaagaatacttcat 964
      |||
Qy 2082 TGCATACGAAGACCATACATAATAACCAACATAACATATACACAAAGAAATACCTTCAT 2141

Db 965 tatttgggaattagtgcatctcaaaagtaatacatatatacaaaactaggcaccacta 1024
      |||
Qy 2142 TATTTGTGGAATTTAGTCGATTTCAAAAGTAATCATATATCAAACTAGGACCACACTA 2201

Db 1025 agtctgattattttgttataatttaataatctcttatgagccctatatattcaaa 1084
      |||
Qy 2202 AGTCTCGATTATTTGTTATTAATTTAATTAATATATCTATGAGCCCTATATATTCAAA 2261

Db 1085 atattgttaacatgaatccatgtttcttttcc 1120
      |||
Qy 2262 ATATTATGTTAAACATGTAATCCATGTGTTCTTTTTC 2297
```

```
RESULT 4
LOCUS Homo sapiens iduronate sulphatase (IDS) gene, complete cds.
DEFINITION L35485
ACCESSION iduronate sulphatase.
KEYWORDS Homo sapiens DNA.
SOURCE Homo sapiens
ORGANISM Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
```

REFERENCE 1 (bases 1 to 36845)
AUTHORS Lu, F., Lu, J., Clingan, R.L., Wentland, M.A., Muzny, D.M., Gu, Y., Nelson, D.L. and Gibbs, R.A.
TITLE Complete DNA sequence of the human iduronate sulphate sulphatase (ids) locus
JOURNAL Unpublished (1994)
COMMENT NCBI gi: 530140
FEATURES
source Location/Qualifiers
1..36845
/organism="Homo sapiens"
/sequenced_mol="DNA"
<1057..1159
/gene="IDS"
/map="Xq27.3-q28"
/number=1
/codon_start=1
Join(1057..1159,1901..2037,2704..2881,5155..5243,7885..8085,9676..9846,15752..15878,19094..19267,22974..23446)
/gene="IDS"
/map="Xq27.3-q28"
/note="NCBI gi: 530141"
/codon_start=1
/product="iduronate sulphate sulphatase"
/translation="MPPRTGRLGLGLVLSVCVALGSETQNSTTDALNVLIVDDLRLGCGDKLVRSPNIDQLASHLLFQNAFAQAVCAPSRVSFLTCRRPDTRL YDENSYRVHAGNESTIPQYFKENGYYTMSGVKVFHPCISSNHTDSDPSYMSFPYP SSEKYENTKCRGDGDLHANILCPVDVLDVPEGLDPKOSTEQAIOLLEKMTSASP FFLAVGYHKPHIPRPYKEFKQLYPLENTLAPDPEVDCGLPPVAYNPWMDIRQEDV QALNI SVYFCIPVDYQRIKIQSYFASVYLDTVQGRLLSALDDIQLANSTIIAFTSD HGNALGEGENAKYSNFQVATHTVPLIFYVPGRTASLPEAGEKLFYLPDPEDSASQME PGRQMDVLVSLFTPLAGLAGLQVPPCPVPFSHVELCREGNLLKHFRRDLEED PYPGNPRELIAYSQYPRSDIPQWNSDKPSIKQIKIMGYSIRTIDYRYTVMVGNPD EFLANFSDIHAGELYFVDSDP LQDRHNYNDSQGGDLFQLMD"
1160..1900
/gene="IDS"
/map="Xq27.3-q28"
/number=1
1901..2037
/gene="IDS"
/map="Xq27.3-q28"
/number=2
/codon_start=1
2038..2703
/gene="IDS"
/map="Xq27.3-q28"
/number=2
2704..2881
/gene="IDS"
/map="Xq27.3-q28"
/number=3
/codon_start=1
2882..5154
/gene="IDS"
/map="Xq27.3-q28"
/number=3
5155..5243
/gene="IDS"
/map="Xq27.3-q28"
/number=4
/codon_start=1
5244..7884
/gene="IDS"
/map="Xq27.3-q28"

repeat_region
6827..7114
/rpt_family="Alu"
repeat_region
7117..7647
/rpt_family="Line1"
exon
7885..8085
/gene="IDS"
/map="Xq27.3-q28"
/number=5
/codon_start=1
8086..9675
/gene="IDS"
/map="Xq27.3-q28"
/number=5
9676..9846
/gene="IDS"
/map="Xq27.3-q28"
/number=6
/codon_start=1
9847..15751
/gene="IDS"
/map="Xq27.3-q28"
/number=6
15752..15878
/gene="IDS"
/map="Xq27.3-q28"
/number=7
/codon_start=1
15879..19093
/gene="IDS"
/map="Xq27.3-q28"
/number=7
19094..19267
/gene="IDS"
/map="Xq27.3-q28"
/number=8
19268..22972
/gene="IDS"
/map="Xq27.3-q28"
/number=8
repeat_region
21573..21864
/rpt_family="Alu"
repeat_region
21867..22051
/rpt_family="Line1"
repeat_region
22098..22267
/rpt_family="Line1"
exon
22974..23446
/gene="IDS"
/map="Xq27.3-q28"
/number=9
25413..25686
/rpt_family="Alu"
repeat_region
26878..27162
/rpt_family="Alu"
repeat_region
29858..30098
/rpt_family="Alu"
repeat_region
30767..31068
/rpt_family="Alu"
repeat_region
33970..34090
/rpt_family="Alu"
BASE COUNT 9117 a 8696 c 8303 g 10729 t
ORIGIN

DB 107; Score 995; Match 100.0%; QryMatch 43.3%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22971 caggcaggaacatccatggaacctgtggaacctgtgtctctttttccacagctggtggac 23030
|||||
Qy 1302 CAGCGAGCAATCCATGGACCTTGTGGAACCTGTCTCTCTTTTCCACAGCTGGCTGGAC 1361
|||||
Db 23031 ttgcaggactgcaggttccacctgcctgccccgttccttcattccacgttgagctgtgca 23090
|||||
Qy 1362 TTGCAGAGCTGACGGTTCCACCTCGCTGCCCGCTTCCTTCATTTCACGTGACGTGCA 1421
|||||
Db 23091 gagaagcaagaacctctggaagcatttcgattccgtgacttgaagagatccgtacc 23150
|||||
Qy 1422 GAGAAGCAAGAAGCTTCTCAAGCAATTTCCGATTCCTGCACTTGGAAAGGATCCGTACC 1481
|||||
Db 23151 tccctggtaatccccgtgaacctgattgctctatagccagtatccccggcttcagacatcc 23210
|||||
Qy 1482 TCCCTGCTAATCCCGTCAACTGATTGCCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
|||||
Db 23211 etcagtggaattctgacaagccgagtttaaaagataaaagatcatgggtattccatc 23270
|||||
Qy 1542 CTCAGTGAATCTGCACGCCGAGTTAAAGATATAAAGATCATGGCTATTCATAC 1601
|||||
Db 23271 gcaaccatagactataggtatctgtgtgggtggcttcaatcctgataaattcttagtca 23330
|||||
Qy 1602 GCACCATAGACTATAGTATGATCTGTGTGGGTGGCTTCAATCTCATGATTTCTAGCTA 1661
|||||
Db 23331 acttttctgaactccatgcagggaactgtatttgggtattctgaacctatccatccaggatc 23390
|||||
Qy 1662 ACTTTTCTGACATCCATGCAGGGAACGTATTTTGTGGAATTCGACCATTCGAGGATC 1721
|||||
Db 23391 acaatatgtataatgattcccaagtggaagatctttccagttgttgatgcttgagttt 23450
|||||
Qy 1722 ACAATATGTAATGATTCCCAAGGTGAGATCTTTTCCAGTTGTTGATGCTTGAGTTT 1781
|||||
Db 23451 tgcacaaccatggatggcaaatgtgatgtccttccctccagctgggtgagagagaggttag 23510
|||||
Qy 1782 TCCCAACCATGATGGCGAATGTGATGTGCTCCCTCCAGCTGCTGAGAGGAGCTTAG 1841
|||||
Db 23511 agctgggtcgtttgtgattaccataaatattggaagcagctgagggctagttaatccaa 23570
|||||
Qy 1842 AGCTGGTGGTTTTGTGATTACCCATAATATTGGAAGCAGCCTCAGGGCTAGTTAATCCAA 1901
|||||
Db 23571 acatgcatcaacaatttggcctgagaatatgttaacagccaaaccttttcgttttagcttt 23630
|||||
Qy 1902 ACATGATCAACAATTTGGCCTGAGATATGTAAAGCAGCAAACTTTTCGTTTGTCTTT 1961
|||||
Db 23631 attaaattataattgtaattggaacctgttttttttaattccctctttttaaac 23690
|||||
Qy 1962 ATTAATAATTATAATGGTAATTGACACAGTTTTTTTTTAATTTCCCTCTCTTTTAAAC 2021
|||||
Db 23691 agttacaggttatttactgaataataacaagcaacaactcaagttagtcataacct 23750
|||||
Qy 2022 AGTTACGGCTTATTTACTGATAAATAACAAGCAACAACTCAAGTTATGTGTATACCTT 2081
|||||
Db 23751 tggatacgaagaccatacataaaccacaataacattatcacaaagaatactttcat 23810
|||||
Qy 2082 TGGATACGAAGACCATACATAATAACAAACATAACATTATACAAAAGATATCTTTTCAAT 2141
|||||
Db 23811 tatttgggaattttagtgatttcaaaagttaataatcatatatacaactaggcaccacata 23870
|||||
Qy 2142 TATTTTGGAAATTTAGTGCATTTTCAAAAAGTAATCATATATCAAACTAGGCACCACTA 2201
|||||
Db 23871 agttcctgattattttgttttaatttaataatataatcttatgagccctatatattcaaa 23930
|||||

Qy 2202 ACTTCTCATTTATTTGTTTATAATTTAAATAATATATCTTATGAGCCCTATATATCAA 2261
|||||

Db 23931 atttatgttaacatgtaatccatgtttcttttc 23965
|||||
Qy 2262 ATATTATGTTAAACATGTAATCCATGTTTCTTTTTC 2296
|||||

RESULT 5
LOCUS H06357 615 bp mRNA EST 21-JUN-1995
DEFINITION Y179J02.r1 Homo sapiens cDNA clone 44414 5' similar to gb:M58342
IDURONATE 2-SULFATASE PRECURSOR (HUMAN);.
ACCESSION H06357
KEYWORDS EST.
SOURCE human clone=44414 library=Soares infant brain INIB vector=Lafmid BA

host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1=Not I
Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTCGAGAAATTCGGCGCCGACGAAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldi.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 615)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lemmon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

GDB: G00-416-955
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 412
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 869909
FEATURES
source
Location/Qualifiers
1..615
/organism="Homo sapiens"
/clone="44414"
/note="human"
BASE COUNT 140 a 155 c 154 g 162 t 4 others
ORIGIN

DB 142; Score 469; Match 93.5%; QryMatch 20.4%; Pred. No. 0.00e+00;
Matches 575; Conservative 0; Mismatches 26; Indels 14; Gaps 14;
Db 1 cttgctgnacaacagagagcactgagcaagccatacagttgttgaaaagatgaaacgctc 60

Db 61 cccctggatgacatcaggcaacggaagacgtccaagccttaaacatcagtgccgta 120
|||||
Qy 919 CCCCTGATGCACATCAGGCAACGGGAAGACGTCCAAGCCTTAAACATCAGCTGCGCGTA 978

Db 121 tggctcaattctctggaatttcacggaaatccggcagagctacttgcctctgtgc 180
|||||
Qy 979 TGGTCCAAATCTCTGGACTTTTCAGGGAATCCGCCAGAGCTACTTTGGCTCTGTCTC 1038

Db 181 atatttgatcacacaggtcgccgcctcttgagtgcttggagcatcttcagctggccaa 240
|||||
Qy 1039 ATATTGGATACACAGTCGCGCGCTCTTGAAGCTTTGGAGCATCTTCAGCTGCCAA 1098

Db 241 cagcaccatcattgatttacctcgatcatgggtgggtcctagtggaacatggagaat 300
|||||
Qy 1099 CAGCACCATTGATTTACCTCGCATCGGTGGCTCTAGGTGAACATGG-AGAAT 1157

Db 301 gggccaaatcacagaatttgatgttctaccatgttccctcgatattctatgttctcg 360
|||||
Qy 1158 GGGCAAATACAGCAATTTTGATGTGTGTACCCATGTTCCCTGTATTTCTATGTTCTGTG 1217

Db 361 gaaggcggcttcaattccggaggcggcgaagaggttttcccttaoctngaccttt 420
|||||
Qy 1218 GAAGACGGCTTCACTTCGCGAGCAGG-CGAGAGCTTTTCCCTTACTCGACCCCTTTT 1276

Db 421 tgattcngcttaacagtgatggagccagggaaggaattccnng 466
|||||
Qy 1277 -GATTCC-GCCTCACAGTTCATGG-AGCCAGGCGCAATCAATG 1319

RESULT 9
LOCUS H05179 469 bp mRNA EST 21-JUN-1995
DEFINITION y185c04.r1 Homo sapiens cDNA clone 45095 5' similar to gb:M58342
IDRONATE 2-SULFATASE PRECURSOR (HUMAN);
ACCESSION H05179
KEYWORDS EST.
SOURCE human clone=45095 library=Soares infant brain 1N1B vector=lafmid BA
host=DH10B (ampicillin resistant) primer=M3R1 Rsite1=Not I
Rsites2-Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5'
AACTCGAAGAAATTCGCGCCGACAGAAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 469)

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennan, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)

COMMENT
GOB: G00-417-636
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence steps: 103
Source: IMAGE Consortium, LINL
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

NCBI gi: 868731
FEATURES
source
Location/Qualifiers
1..469
/organism="Homo sapiens"
/clone="45095"
/note="human"
BASE COUNT 102 a 128 c 119 g 116 t 4 others
ORIGIN

DB 142; Score 412; Match 96.6%; QryMatch 17.9%; Pred. No. 0.00e+00;
Matches 450; Conservative 0; Mismatches 11; Indels 5; Gaps 5;
Db 1 gaacatcacctggccccgatcccgaggtccctgatggctaccacctggcctacaa 60
|||||
Qy 859 GAACATCACCTGGCCCCGATCCGAGTCCCTGATGGCTTACCCCTGTGGCTTACAA 918

Db 61 cccctggatgacatcaggcaacggaagacgtccaagccttaaacatcagtgccgta 120
|||||
Qy 919 CCCCTGATGCACATCAGGCAACGGGAAGACGTCCAAGCCTTAAACATCAGTGTGCGGTA 978

Db 121 tggctcaattctctggaatttcacggaaatccggcagagctacttgcctctgtgc 180
|||||
Qy 979 TGGTCCAAATCTCTGGACTTTTCAGCGGAAAATCCGCCAGAGCTACTTTGGCTCTGTCTC 1038

Db 181 atatttgatcacacaggtcgccgcctcttgagtgcttggagcatcttcagctggccaa 240
|||||
Qy 1039 ATATTGGATACACAGTTCGCGCGCTCTTGAAGCTTTTCCCTTACTCGACCCCTTTT 1098

Db 241 cagcaccatcattgatttacctcgatcatgggtgggtcctagtggaacatggagaat 300
|||||
Qy 1099 CAGCACCATTGATTTACCTCGCATCGGTGGCTCTAGGTGAACATGG-AGAAT 1157

Db 301 gggccaaatcacagaatttgatgttctaccatgttccctcgatattctatgttctcg 360
|||||
Qy 1158 GGGCAAATACAGCAATTTTGATGTGTGTACCCATGTTCCCTGTATTTCTATGTTCTGTG 1217

Db 361 gaaggcggcttcaattccggaggcggcgaagaggttttcccttaoctngacctttt 420
|||||
Qy 1218 GAAGACGGCTTCACTTCGCGAGCAGG-CGAGAGCTTTTCCCTTACTCGACCCCTTTT 1276

Db 421 tgattcngcttaacagtgatggagccagggaaggaattccnng 466
|||||
Qy 1277 -GATTCC-GCCTCACAGTTCATGG-AGCCAGGCGCAATCAATG 1319

RESULT 10
LOCUS T08577 378 bp mRNA EST 03-AUG-1993
DEFINITION EST06469 Homo sapiens cDNA clone HIBG50 5' end similar to
Iduronate 2-sulfatase.
ACCESSION T08577
KEYWORDS EST.
SOURCE Human clone=HIBG50 library=Infant brain, Bento Soares primer=M13
Reverse.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 378)
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
Nature Genet. 4, 373-380 (1993)

CONTACT: ADAMS, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@igr.org.

NCBI gi: 389605
Location/Qualifiers
source
1..378
/organism="Homo sapiens"
/clone="H18G50"
/note="Human"

BASE COUNT 129 a 80 c 70 g 99 t
ORIGIN

DB 56; Score 369; Match 99.5%; QryMatch 16.1%; Pred. No. 0.00e+00;
Matches 376; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 2 gccgtactgttttaaaagagggaattaaaaaaa-cgtgtccaattaccaattata 60
|||||
|||||

Cp 2030 GCCGTAAGTGTTTAAAGAGAGGAATTAAGGAACTGGTCCATTACCAATTATA 1971
|||||

Db 61 aattttaagaactaaacgaagggttgctgtgtacatatcttcaggccaaattgtt 120
|||||
|||||

Cp 1970 AATTTTAAGAAGCTAAAGCAAGGTTGGCTGTACATATTCAGGCGCAATTGTT 1911
|||||

Db 121 gatcatgtttgatttaactagccctcaggctgttccaatatattggttaacacaaa 180
|||||
|||||

Cp 1910 GATGCATGTTGGATTAACTAGCCCTCAGGCTGCTTCCAATATTATGGGTAAATCAAAA 1851
|||||

Db 181 cagccagcttaactctctctcaccagctggaaggagcacatcaatttgcaccca 240
|||||
|||||

Cp 1850 CGACCAGCTTAACTCCTCTCTCACCAGCTGGAAGGAGCACATCACATTTGCCATCCA 1791
|||||

Db 241 tggttgcaaaactcaaggcatcaacaactggaagaatctccaccttgggaatcattat 300
|||||
|||||

Cp 1790 TGGTTGGCAAACTCAGGCGTCACAACTGGAAAGATCTCCACCTTGGGATCATAT 1731
|||||

Db 301 acatattgtgatctgcaatgggtgcagaatccacaaaatacagttccctgcatgggtgt 360
|||||
|||||

Cp 1730 ACAATTGTGATCTGCAATFGGCTGAGATCCACAAATATACGTTCCCTCGATGATGT 1671
|||||

Db 361 cagaaaagttagctagaa 378
|||||

Cp 1670 CAGAAAGTTAGCTAGAA 1653

RESULT 11
LOCUS R15374 484 bp mRNA EST 13-APR-1995
DEFINITION yf90e05.r1 Homo sapiens cDNA clone 29589 5' similar to gb:M58342
IDUNONATE 2-SULFATASE PRECURSOR (HUMAN);
ACCESSION R15374

KEYWORDS
SOURCE human clone=29589 library=Soares infant brain 1NTB vector=Iafmid BA
host=DH10B (ampicillin resistant) primer=M3RPI Reitel=Not I
Reite2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5']

AACTGAAGAATTCGGCGCCGAGGAATTTTFTTTTTTTT 3'; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Iafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 484)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

GDB: G00-401-936
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 350
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 769647
Location/Qualifiers
source
1..484
/organism="Homo sapiens"
/clone="29589"
/note="human"

BASE COUNT 112 a 128 c 121 g 114 t 9 others
ORIGIN

DB 38; Score 360; Match 97.8%; QryMatch 15.7%; Pred. No. 0.00e+00;
Matches 402; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

Db 1 ctgtcngagcaaacagagcaactgagcaagccatacagttgttgaaagatgaaacgt 60
|||||
|||||

Qy 709 CTTGCTGCA-CAACAGAGCACTGAGCAAGCCATACAGTTCTTGAAAGATGAAACGT 767
|||||

Db 61 cagcagctctcttcttctgtggtgtatcataagccacacatcccccttcagatacc 120
|||||
|||||

Qy 768 CAGCAGCTCCTTTCTCTGCGCTGGGTATCATAGCCACACATCCCTTCAGATACC 827
|||||

Db 121 ccaaggaattcagaagttgtatcccttgagaaacatcacctggcccccgatcccgagg 180
|||||
|||||

Qy 828 CAGAGAAATTCAGAAATTGTATCCCTTGAGAGAAATCACCTGGCCCCCGATCCCGAGG 887
|||||

Db 181 tccctgatggcctaccctctggcctacaaccctggatggagcatcaggcaacgggaag 240
|||||
|||||

Qy 888 TCCTCATGGCCCTTACCCTCTGGCCCTACAAACCCCTGGATGGACATCAGGCAACGGGAAG 947
|||||

Db 241 aggtccagccttaaacatcagtggtgcgtatgttcccaattctctgtgactttcagcgga 300
|||||
|||||

Qy 948 ACCTCAAGCTTTAAACATCAGTGTCCGTATGTCCTGTCCTGTCGATTTTCAGCGCA 1007
|||||

Db 301 aaatccgcagagagcttaacttggctctgtgttcatttggatacacacaggtcgccgct 360
|||||

Qy 1008 AATCCGCCAGAGCT-ACITTCCTCTGTCT-CATATTTGGATACACAGCTGGCGCGCT 1065
|||||
Db 361 cttggagctcttgggacgatcttcctcagctgggccaacagcaccatcatg 411
|||||
Qy 1066 CTTG-AGTCTTTGG-ACGATCTT-CAGCTGG-CCACAGCAGCATCATG 1112
|||||
RESULT 12
LOCUS H06300 439 bp mRNA EST 21-JUN-1995
DEFINITION y179h02.s1 Homo sapiens cDNA clone 44414 3' similar to gb:M58342
IDURONATE 2-SULFATASE PRECURSOR (HUMAN);
ACCESSION H06300
KEYWORDS EST.
SOURCE human clone=44414 library=Soares infant brain LNB vector=Lafmid BA
host=DH10B (ampicillin resistant) primer=Promega -2lml3 Reitel=Not
I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAGATCCGCCCGCCGACGAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT GDB: G00-416-955
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 315
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 869852
Location/Qualifiers
source 1..439
/organism="Homo sapiens"
/clone="44414"
/note="human"

BASE COUNT 135 a 83 c 83 g 133 t 5 others

ORIGIN

DB 142; Score 332; Match 96.1%; QryMatch 14.5%; Pred. No. 0.00e+00;
Matches 398; Conservative 0; Mismatches 5; Indels 11; Gaps 11;
Db 17 tttgttcttggattatttcagtaataagcgcgaactnttttttaaaagagggaatt 76

Cp 2061 TTGTGTTGCTTTGTATTATTATTCAGTAAATAGCGTAACCTCTTTTAAAAAGACGGAAAT 2002
|||||
Db 77 aaaaaaaactggtccattaccattataaatttttaataaagactaaacgaaggtt 136
|||||
Cp 2001 AAAAAAAGAACTGGTCCAAATACCAATTTATAAATTTTAAATAAGACTAAACGAAGGCT 1942
|||||
Db 137 tggctgttacattctcagggccaaattnttgatgatgttgggattactagccctcag 196
|||||
Cp 1941 TGGCTGTATACATATCTTCAGGCCAAATGTTGATGCATGTTTGGATTAACTAGCCCTCAG 1882
|||||
Db 197 gctgactccaataatattgggtaatacacaagaccagctcttaactcctctcca 256
|||||
Cp 1881 GCTG-CTTCCAAATATTATGGTAATCAAAAACGACGCTTAACCTCTCTCAACA 1823
|||||
Db 257 gctggaaggagcacatcacattgcccattccatgggttgggcaaaactcaagggaatcaa 316
|||||
Cp 1822 GCTGGAAGGAGCAGCATCACATTTGCCATCCATGG-TTGG-CAAAACTCAAGG-CATCAA 1766
|||||
Db 317 caactgggaaaaggatctccacttgggggaatcattacatatgtgatccctggaatg 376
|||||
Cp 1765 CAACTGG-AAAAG-ATCTCCACCTTGGG-AATCATTTATACATATTGTGATCCTG-CAATG 1710
|||||
Db 377 gggtcagantccacaanaatacagttccctggcagtggtgntgtcagaaaagttag 430
|||||
Cp 1709 GG-TCAGATCCACAAAATAACAGTTCCTG-CATGG-ATGTCAGAAAAGTTAG 1659
|||||

RESULT 13
ID HS300154 standard; RNA; EST; 439 BP.
AC H06300;
DT 23-JUN-1995 (Rel. 44, Created)
DT 23-JUN-1995 (Rel. 44, Last updated, Version 1)
DE y179h02.s1 Homo sapiens cDNA clone 44414 3' similar to gb:M58342
DE IDURONATE 2-SULFATASE PRECURSOR (HUMAN);
KW EST.
OS Homo sapiens (human)
OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;
OC Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
RN [1]
RP 1-439
RA Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
RA Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
RA Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
RA Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., Wilson, R.;
RT "The WashU-Merck EST Project";
RL Unpublished.
CC GDB: G00-416-955 Contact: Wilson RK WashU-Merck EST Project
CC Washington University School of Medicine 4444 Forest Park Parkway,
CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810
CC Email: est@watson.wustl.edu High quality sequence stops: 315
CC Source: IMAGE Consortium, LNL This clone is available royalty-free
CC through LNL; contact the IMAGE Consortium (info@image.llnl.gov)
CC for further information. NCBI gi: 869852
FH Key
FH Location/Qualifiers
FT source 1..439
FT /organism="Homo sapiens"
FT /clone="44414"
FT /note="human"

SQ Sequence 439 BP; 135 A; 83 C; 83 G; 133 T; 5 other;
DB 3; Score 332; Match 96.1%; QryMatch 14.5%; Pred. No. 0.00e+00;
Matches 398; Conservative 0; Mismatches 5; Indels 11; Gaps 11;

Db 17 ttgtgtgtgttatttattcagtaataagcgttaactntttttaaagaggggaatt 76
|||||
Cp 2061 TTTGTTTGTCTTGTATTATTCAGTAATAGCCGTAACTGTTTTTAAAGACGGGAAT 2002
|||||
Db 77 aaaaaaaactggtccaattaccattataaatttttaataagagactaaacgaaaaagtt 136
|||||
Cp 2001 AAAAAAAACTGGTCCAATTACCAATTTAAATTTTAAATTAAGACTAAACGAAAGTT 1942
|||||
Db 137 tggctgttaacattctcagggccaaattnttgatgcatgtttgttgatttaactagccctcag 196
|||||
Cp 1941 TGGCTGTTCATATTCCTCAGGCCAAATTTGTTGATGATGTTTGGATTAACTAGCCCTCAG 1882
|||||
Db 197 gctgacttccaattatggtgaatcacaaaacgaccagctctaaactcctctcaca 256
|||||
Cp 1881 GCTG-CTTCATATATATGGGTAAATCAGAAACGACGAGCTTAACTCTCTCTCACC 1823
|||||
Db 257 gctgaaaggagcacatcacatttgcattcctggttggtggcaaaactcaagggcatcaa 316
|||||
Cp 1822 GCTGAAGGAGCACAATCAATTCGATCCATGG-TTGG-CAAAACTCAAGG-CATCAA 1766
|||||
Db 317 caactggaaaaggatctccacttgggggaatcattatcatatgtatctcggcaatg 376
|||||
Cp 1765 CAACTGG-AAAAG-ATCTCCACCTTGGG-AATCAITATACATATTGTCATCTG-CAATG 1710
|||||
Db 377 gggtcagantccacaanaatcacagttccctcggcatggngtgcagaaaagttag 430
|||||
Cp 1709 GG-TCAGAAATCCACAAATACAGTTCCCTCG-CATGG-ATGTCAGAAAGCTTAG 1659
|||||

RESULT 14

LOCUS HSC0SC011 327 bp RNA EST 14-FEB-1995

DEFINITION H. sapiens partial cDNA sequence; clone c-0ac01.

ACCESSION F05959

KEYWORDS partial cDNA sequence; transcribed sequence fragment.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Carnivora; Hominidae; Homo.

1 (bases 1 to 327)

Genexpress.

Direct Submission

Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.

Genethon, B.P. 60, 91002 Evry Cedex France and Genetique

Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801

Villejuif Cedex France. E-mail: genexpress@genethon.fr

2 (bases 1 to 327)

Genexpress.

The Genexpress cDNA program

Unpublished

3 (bases 1 to 327)

Auffray, C., Behar, G., Bois, F., Boucher, C., da Silva, C.,

Devignes, M.D., Duprat, S., Houllgatte, R., Jumeau, M.N., Lamy, B.,

Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,

Sebastien, K., Kabbaj, C. and Tessier, A.

INAGS: Integrated molecular analysis of the human genome and its

expression

C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)

Cloning method: total mRNA was oligo-(dT) primed and directionally

cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA

vector;

Sequencing method: single read, full automatic;

Primer: M13 reverse

cDNA sequence colinear to mRNA

Stretch removed: nothing

Normalization method: Bento Soares, P.N.A.S. 91:9228-9232 (1994);

Genexpress library_idt: C;

Genexpress_sequence_idt: ylc-0sc01.

NCBI gi: 669775

location/Qualifiers

source

1..327

/organism="Homo sapiens"

/clone_lib="normalized infant brain cDNA from B. Soares,

Psychiatry Dept. Columbia University USA"

/sex="female"

/tissue_type="total brain"

/dev_stage="3 months old"

/isolates="muscular atrophy patient"

BASE COUNT 75 a 95 c 75 g 78 t 4 others

ORIGIN

DB 26; Score 321; Match 98.5%; QryMatch 14.0%; Pred. No. 0.00e+00;

Matches 322; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 aaaaatgaaacgtcagccagcttnttctgcccgttggtgatcataagccacaca 60

Qy 753 AAAGATCAAAACGCTCAGCCAGTCTCTTTCTTCGCGCTGGGTATCATAGCCACACA 812

Db 61 tcccttcagatatacccaaggacttcagaagttgtatcccttgaggagaacatcacctgg 120

Qy 813 TCCCTTCAGATACCCCAAGAAATTCAGAAAGTTGTATCCCTTGGAAGACATCACCTGG 872

Db 121 ccccgatcccgaggttcctgacccctaccctctggtccctacacccctggatggaca 180

Qy 873 CCGCGATCCCGAGGTCTCTGATCCCTTACCCCTTGGCTTACACCCCTGGATGGACA 932

Db 181 tcaggcaacgggaagcagtcacagccttaaacatcagtgccgtatggtccaatcctg 240

Qy 933 TCAGGCAACGGGACAGCTCCAGGCTTAAACATCAGTGTGCGGTATGGTCCAAATTCCTG 992

Db 241 tggacttcagcgagaaatccgncagagactacttgctctctgtgcatatttgatacac 300

Qy 993 TGGACTTTACGGGAAATCCCGAGACTACTTTGGCTGTCTCATATTGGATACAC 1052

Db 301 aggtcgccgctctcttgagtgcttgg 327

Qy 1053 AGGTGCGCGCCCTCTTGGTCTTTGG 1079

RESULT 15

LOCUS

T71926

423 bp

mRNA

EST

16-MAR-1995

DEFINITION

ye07a10.r1 Homo sapiens cDNA clone 117018 5'

IDURONATE 2-SULFATASE PRECURSOR (HUMAN);

T71926

ACCESSION

EST.

KEYWORDS

SOURCE

human clone=117018 library=Stratagene lung (#937210)

vector=pBluescript SK- host=SOLR cells (kanamycin resistant)

primer=M13RPI Reitel=EcoRI Reitel=XhoI Normal lung tissue from a 72

year old male. Cloned unidirectionally. Primer: Oligo dT. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:

5'-GAATTCGGCAGAG-3'; 3' adaptor sequence:

5'-CTCGAGTTTTTTTTTTTTTTTTTTT-3'.

Homo sapiens

Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

REFERENCE

AUTHORS

Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 423)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE

The WashU-Merck EST Project

Unpublished (1995)

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 294
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 686447

Location/Qualifiers

source 1..423

/organism="Homo sapiens"

/clone="117018"

/notes="human"

BASE COUNT 88 a 114 c 101 g 119 t 1 others

ORIGIN

DB 71; Score 319; Match 95.8%; QryMatch 13.9%; Pred. No. 0.00e+00;
Matches 406; Conservative 0; Mismatches 4; Indels 14; Gaps 14;

Db 1 attttgatgtacctaccctgtccctgatattctctgtcttctggaagcgggttcac 60

|||||

Qy 1173 ATTTTGATGTTCTACCCATGTTCCCTGATATTCTATGTTCTGGAAGCAGCGCTTCAC 1232

Db 61 ttccggag-caggcgagaagctttcccttaoctgcaccccttttgattccgcctcacagt 119

|||||

Qy 1233 TTCCGGAGCAGCGCAGACCTTTTCCTTTACTCGACCCCTTTTGATTCCGCCCTCACAGT 1292

Db 120 tgatggagccaggcaggcaatccatgacctgtggaactgtgtctctttttcccaagc 179

|||||

Qy 1293 TGATGGAGCAGCGCAGCAATCCATGACCTTGTGGAACTTGTTGTCCTTTTCCACGC 1352

Db 180 ttgctgacctgcaggactgcaggttccacctgcctgcctgcccccttccttcatttcacgttg 239

|||||

Qy 1353 TGGCTGCACTTCGAGACTCGACGTTCCACCTGCGTGCCTGCTTCCTTCATTTCACGTTG 1412

Db 240 agctgtgcaggaagcagaaccttcttggagcattttcgattccggaacttgggaag 299

|||||

Qy 1413 ACCTGTCGAG-AGAAGGCAAGAACCTCTC-AAGCATTTTCGATTCCGTCACTTGG-AAG 1469

Db 300 aggatccgtacctccctgggtaaatcccccggtggaacttgattgctatagccagatgc 359

|||||

Qy 1470 AGGATCGTACCTCCCTGG-TAA-TCCCC-GTG-AACT-GATT-GCCTATAGCCAGTATC 1523

Db 360 cccgggcttcagacatccctcagtggaatnttgacaagccgaggttttaaaagatata 419

|||||

Qy 1524 CCCC-GCTTCAGACATCCCTCAGTGG-AATTCT-GACAGCCGAGTTT-AAAAGATATA 1579

Db 420 aaga 423

||||

Qy 1580 AAGA 1583

Search completed: Wed Oct 4 14:43:28 1995
Job time : 3284 secs.

WAPSEH (TH)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Oct 4 13:42:53 1995; MasPar time 126.71 Seconds
925.108 Million cell updates/sec

Tabular output not generated.

Title: >US-08-345-212-1
Description: (1:2297) from US08345212.seq

Perfect Score: 2297
N.A. Sequence: 1 CGGCTGTGTGCGCAGCTTT.....TAATCCATGTTTCTTTTCC 2297
Comp: GCGCAGACAGCGCTCAGAA.....ATTAGTACAAAGAAAGG

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-geneseq
1 n-gen1
2 n-gen2
3 n-gen3
4 n-gen4
5 n-gen5
6 n-gen6
7 n-gen7
8 n-gen8
9 n-gen9
10 n-gen10
11 n-gen11

Statistics: Mean 10.033; Variance 5.812; scale 1.726

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
c 1	231	10.1	233	10	Q77329 Human genome fragment	4.48e-140
c 2	99	4.3	1047	2	Q10572 Human Natriuretic Pep	4.33e-47
c 3	72	3.1	1047	2	Q10572 Human Natriuretic Pep	4.36e-29
c 4	48	2.1	204	1	N81164 Base substituted E.co	6.20e-14

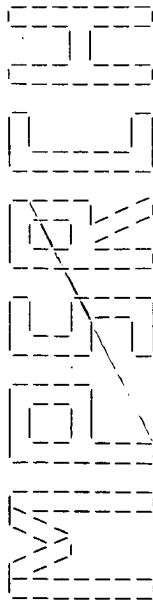
5	48	2.1	204	1	N81164	Base substituted E.co	6.20e-14
6	45	2.0	91	9	Q51746	Oligonucleotide probe	3.90e-12
c 7	40	1.7	91	9	Q51746	Oligonucleotide probe	3.30e-09
8	37	1.6	501	3	N50033	Sequence encoding new	1.68e-07
9	35	1.5	501	3	N50023	Sequence encoding new	2.19e-06
10	35	1.5	501	3	N50031	Sequence encoding new	2.19e-06
11	35	1.5	501	3	N50025	Sequence encoding new	2.19e-06
12	35	1.5	501	3	N50027	Sequence encoding new	2.19e-06
13	35	1.5	501	3	N50028	Sequence encoding new	2.19e-06
14	35	1.5	501	3	N50032	Sequence encoding new	2.19e-06
15	35	1.5	498	3	N50034	Sequence encoding new	2.19e-06
16	33	1.4	501	3	N50029	Sequence encoding new	2.70e-05
17	33	1.4	501	3	N50030	Sequence encoding new	2.70e-05
18	31	1.3	501	3	N50024	Sequence encoding new	3.15e-04
19	30	1.3	565	6	Q35072	HCV envelope region n	1.05e-03
20	30	1.3	501	3	N50026	Sequence encoding new	1.05e-03
21	25	1.1	3871	2	N71302	HSV-1 gB and surround	3.29e-01
22	25	1.1	39	7	Q51787	Mixed oligonucleotide	3.29e-01
c 23	24	1.0	5393	5	Q34611	Plasmid pR3.	9.71e-01
c 24	23	1.0	501	3	N50031	Sequence encoding new	2.80e+00
c 25	23	1.0	501	3	N50023	Sequence encoding new	2.80e+00
c 26	23	1.0	501	3	N50032	Sequence encoding new	2.80e+00
c 27	23	1.0	501	3	N50029	Sequence encoding new	2.80e+00
c 28	23	1.0	498	3	N50034	Sequence encoding new	2.80e+00
c 29	23	1.0	1950	1	Q04472	Human papilloma virus	2.80e+00
c 30	23	1.0	501	3	N50028	Sequence encoding new	2.80e+00
c 31	22	1.0	3191	1	Q04092	3.2 kb KpnI-PvuII fra	2.80e+00
c 32	22	1.0	7241	3	Q15140	Genomic clone LE-ACC2	7.82e+00
c 33	22	1.0	36	2	Q11195	Ballast Constituent c	7.82e+00
c 34	22	1.0	3582	8	Q50879	HsmGluRI DNA.	7.82e+00
c 35	22	1.0	39	7	Q51787	Mixed oligonucleotide	7.82e+00
c 36	22	1.0	363	8	Q59724	Human brain Expressed	7.82e+00
37	22	1.0	8920	11	Q62924	Carbamoyl-phosphate-s	7.82e+00
38	22	1.0	2504	6	Q36467	M. scrofulaceum shutt	7.82e+00
c 39	22	1.0	565	6	Q35072	HCV envelope region n	7.82e+00
c 40	21	0.9	4221	11	Q64046	A253 derived bcl-I cD	2.12e+01
c 41	21	0.9	2463	7	Q43711	Sequence of the human	2.12e+01
c 42	21	0.9	5852	2	Q11710	Dictyostelium plasmid	2.12e+01
c 43	21	0.9	48	9	Q53202	CYCD1-H034 first poly	2.12e+01
c 44	21	0.9	48	9	Q53204	CYCD1-G068 first poly	2.12e+01
c 45	21	0.9	501	3	N50024	Sequence encoding new	2.12e+01

ALIGNMENTS

RESULT 1
ID Q77329 standard; DNA; 233 BP.
AC Q77329;
DT 15-NOV-1994 (first entry)
DE Human genome fragment (Preferred).
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN M09401548-A.
PD 20-JAN-1994.
PR 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 470; 616pp; English.

CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
 CC the placenta or bone marrow comprise any of: (A) a sequence
 CC selected from (Q76401-Q77613), (B) an allelic variation of a
 CC sequence as described in (A), or (C) a sequence complementary
 CC to (A) or (B).
 CC Preferred sequences exhibit no more than 90% homology to a human
 CC sequence known per se.
 SQ Sequence 233 BP; 77 A; 22 C; 38 G; 96 T;
 DB 10; Score 231; Match 99.6%; QryMatch 10.1%; Pred. No. 4.48e-140;
 Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1 ttaacataattttgaatataatagagctcataagatatattattataataataa 60
 Cp 2273 TTAACATAATATTTGAATATATAGGCTCATAGATATATATTAAATATAACAAA 2214
 Db 61 taatcagaacttagtggtgcttagttgatatatgatacttttgaatgcaactaa 120
 Cp 2213 TAATCAGAACTTAGTGTGCTAGTTTGCATATATGATTACTTTTGAATGCACTAA 2154
 Db 121 attccacaataagaagtattcttctgtatataatgttggttattgtattgtatgg 180
 Cp 2153 ATTCCACAATAATGAAGTATCTTTGCTATATGTTATGTTGGTATTATGATGG 2094
 Db 181 tcttcgtatccaaagctatgacataacttgattgtttgttctgtgattatt 233
 Cp 2093 TCTTCGTATCCAAAGCTATGACAACTTGCTGTTGCTTGTGTTGTTATTATT 2041
 RESULT 2
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DT 09-APR-1991 (first entry)
 DE Human Natriuretic Peptide Receptor B.
 KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
 KW hyperaldosteronism; glaucoma; guanyl cyclase.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal sequence
 FT Protein 12
 FT /label= mature NPRB
 FT Domain 23..455
 FT /label= extracellular domain
 FT /note= "binds natriuretic peptides A,B and C]"
 FT Domain 456..456
 FT /label= transmembrane domain
 FT Domain 479..1047
 FT /label= cytoplasmic domain
 FT /note= "GC and protein kinase activity"
 FT Modified-site 24..26
 FT /label= N-glycosite
 FT Modified-site 35..37
 FT /label= N-glycosite
 FT Modified-site 161..163
 FT /label= N-glycosite
 FT Modified-site 195..197
 FT /label= N-glycosite
 FT Modified-site 244..246
 FT /label= N-glycosite
 FT Modified-site 277..279
 FT /label= N-glycosite
 FT Modified-site 349..351
 FT /label= N-glycosite

FT Modified-site 600..602
 FT /label= N-glycosite
 PN W09100292-A.
 PD 10-JUN-1991.
 PF 22-JUN-1990; 003586.
 PR 23-JUN-1989; US-370673.
 PA (GETH) GENENTECH INC.
 PI Chang M, Goeddel D, Lowe D;
 DR MPI; 91-036711/05.
 DR N-PSDB; Q10324.
 FT Natriuretic protein receptor B - for diagnosis and treatment of
 PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
 PS Claim 3; Fig 1; 49pp; English.
 CC The sequence was derived from the DNA encoding natriuretic peptide
 CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
 CC kinase activity. The DNA can be inserted into expression vectors
 CC for the prodn. of the protein, opt. after being mutated to produce
 CC NPRB analogues. The protein has a mol wt. of 115 KD (calculated Mr=
 CC 114,952). The protein (or variants) can be used in treatment of
 CC natriuretic peptide disorders, and also to isolate peptides using
 CC affinity chromatography. Antibodies with affinity for NPRB can
 CC also be prepd.
 SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;
 DB 2; Score 99; Match 9.1%; QryMatch 4.3%; Pred. No. 4.33e-47;
 Matches 79; Conservative 243; Mismatches 537; Indels 7; Gaps 7;
 Db 86 dhknyhdndnnngcvcynaasvarnashrnntnntagavasneakndhytrnvtgns 145
 Cp 2050 TGTATTTTTCAGTAAT-AGCGCTAACTGTTTAAAGAGCGGAATTAATAAAAAA 1992
 Db 146 ankgnnvnvtnhqnwtataannynndartddrlnhntngvnnnngsnnsnvhnvvar 205
 Cp 1991 CTGCTCCCAATACCAATATATAATTTTAAAGAGCTAAACGAAAGGTTGGCTGTAC 1932
 Db 206 nnggnnathnrangrnvcgnnnmnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 265
 Cp 1931 AVATTCTCAGGCAAAATTTGATGATGTTTGGATTAACTAGCCCTCAGCTGCTTCCA 1872
 Db 266 ntratrgrnwnhtrmananranntvntyrnnnnnnnnnnnnnnnnnnnnnnnnnnnn 325
 Cp 1871 ATATTATGGGTAATCACAAGACGACC-AGCTCTAATCTCTCTCACCAGCTGGAAGG 1813
 Db 326 gnsnnnnnagcnydgnnnyanvnnntnnggtrndgrnrvnkmgrryhgvtgmvmdkn 385
 Cp 1812 AGCAGATCACATTTCCCATCCATGCTGGCAAACTCAAGGCATCAACAACTGGAAAGA 1753
 Db 386 ndrntdvnwagndsdgdnnaahysganknnwtgrnnnnwkvqansdnncanddd 445
 Cp 1752 TCTCCACCTTGGCAATCATTTATACATATTGCTGCTCTCCTCAATGGGTGCAATCCACAAA 1693
 Db 446 nscdktnstnanvngtntnnngvsnnnnrknnnnknnasmmwrrnrvnnnnnnnn 505
 Cp 1692 TAGACTTCCCTGCGATGTCAGAAAAGTTAGCTAGAAATTCATCAGGATTTGAAGCCA 1633
 Db 506 ryhkgagrrntnngssysgmtahgkynnnantghnkvnnvankhvknnrnntrnn 565
 Cp 1632 ACCACACACTACTTACTATGCTGCTGCTATGCGATGCGATAGCCCATGATCTTATATCT 1573
 Db 566 nnnkhmrdivnnhntnrrngacdnnnncnvtncyrsgmndnnnnndnnnnnnnnnn 625
 Cp 1572 TTTAAACTCGGCTTGTGAGAAATTCACACTGAGGATGCTGAGCGCGGGGATAC-TGGCT 1514
 Db 626 dnvkgmannhnsnshgsknsncvvdsvrnnvknktdygnasnrstannndnnnyaknn 685



(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 4 11:58:26 1995; MasPar time 7.68 Seconds
454.954 Million cell updates/sec

Tabular output not generated.

Title: >US-08-345-212-2

Description: (1:550) from US08345212.pep

Perfect Score: 4069

Sequence: 1 MPPRTGRGLMLGLVLSV.....QDHNNYDSQGDLFQLLMP 550

Scoring table: PAM 150

Gap 11

Searched: 53402 seqs, 6354270 residues

Database: a-geneseq

1 a-gen1
2 a-gen2
3 a-gen3
4 a-gen4
5 a-gen5
6 a-gen6
7 a-gen7
8 a-gen8
9 a-gen9
10 a-gen10

Statistics: Mean 36.792; Variance 159.943; scale 0.230

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	ID	Description	Pred. No.
1	172	4.2	556	9 R51355	Sequence of murine bo	2.75e-05
2	110	2.7	1671	1 R06341	Signal peptide and fi	1.74e+00
3	110	2.7	2721	2 P70647	Sequence of N-termina	1.74e+00
4	107	2.6	351	8 R44211	Caffeine demethylase.	2.85e+00
5	97	2.4	1301	5 R28304	Acid alpha-amylase en	1.42e+01
6	96	2.4	1074	4 R24102	Marek's disease virus	1.66e+01
7	92	2.3	817	9 R47201	DPM2 mannosyl transfe	3.09e+01
8	90	2.2	455	7 R39344	EpiC protein.	4.20e+01

9	90	2.2	969	8 R41662	Paired basic amino ac	4.20e+01
10	90	2.2	490	8 R41663	Paired basic amino ac	4.20e+01
11	87	2.1	112	6 R30872	Prod. of cytochrome p	6.61e+01
12	86	2.1	329	10 R51284	L-fucose dehydrogenas	7.68e+01
13	85	2.1	245	9 R47063	Protein allergen of D	8.90e+01
14	85	2.1	222	9 R52742	Protein allergen of D	8.90e+01
15	85	2.1	320	9 R49920	Protein allergen of D	8.90e+01
16	85	2.1	245	10 R51727	Der p I.	8.90e+01
17	85	2.1	499	2 R11616	Modified human hepari	8.90e+01
18	85	2.1	245	7 R39359	Dermatophagoides pter	8.90e+01
19	85	2.1	245	4 R22433	Der p I allergen.	8.90e+01
20	85	2.1	245	1 P94864	Pre pro Der p I from	8.90e+01
21	84	2.1	3712	3 R13896	ACV synthetase.	1.03e+02
22	84	2.1	2749	3 R13887	Inositol-3-phosphate	1.03e+02
23	84	2.1	842	4 P93712	Sequence of infectio	1.03e+02
24	84	2.1	480	3 P60263	Sequence of new prote	1.03e+02
25	84	2.1	3639	8 R40227	ACVS.	1.03e+02
26	84	2.1	521	8 R47068	Mammalian chromaffin	1.03e+02
27	83	2.0	153	2 P80665	Sequence encoded Haem	1.20e+02
28	83	2.0	270	3 P60130	Murine IL-1 precursor	1.20e+02
29	82	2.0	271	3 P60279	Sequence of mature hu	1.38e+02
30	82	2.0	271	2 R06149	Sequence of human int	1.38e+02
31	82	2.0	271	1 P90421	Human interleukin 1	1.38e+02
32	82	2.0	900	1 R06545	Viral enhancing facto	1.38e+02
33	82	2.0	602	7 R37442	Full-length human pse	1.38e+02
34	82	2.0	98	6 R34275	Human TNF binding ant	1.38e+02
35	82	2.0	271	3 P60129	Human IL-1 precursor.	1.38e+02
36	82	2.0	220	8 R47340	Fragment of chromaffi	1.38e+02
37	82	2.0	271	1 P90108	Human interleukin-1 a	1.38e+02
38	82	2.0	533	6 R37585	E.coli malate synthas	1.38e+02
39	82	2.0	271	1 P90535	Human interleukin-1 (1.38e+02
40	82	2.0	113	6 R34278	Human Vlbmbal-1-1 li	1.38e+02
41	81	2.0	410	9 R54663	Transcription factor	1.60e+02
42	81	2.0	406	10 R57030	Rat 5-HT4 receptor en	1.60e+02
43	81	2.0	964	3 R13618	C3 vegetable PEPC.	1.60e+02
44	81	2.0	738	3 R13251	PECAM-1.	1.60e+02
45	81	2.0	111	2 R12263	Anti-human Rhd FOG-B	1.60e+02

ALIGNMENTS

RESULT 1
ID R51355 standard; Protein; 556 AA.
AC R51355;
DT 12-OCT-1994 (first entry)
DE Sequence of murine bone-related sulphatase-like precursor
DE protein OSF-8.
KW OSF-8; bone-related sulphatase-like protein; osteoporosis;
KW Paget's disease; osteomalacia; therapy; diagnosis.
OS Mus musculus.
PN AU9344921-A.
PD 03-MAR-1994.
PF 27-AUG-1993; 044921.
PR 28-AUG-1992; JP-230030.
PR 03-DEC-1992; JP-324034.
PA (FARH) HOECHST JAPAN LTD.
PA (FARH) HOECHST JAPAN KK.
PI Amann E, Ito T, Otawara-hamamoto Y, Takeshita S;
DR MPI; 94-109856/14.
DR N-PSDB; Q51355.
PT Bone-related sulphatase-like protein, OSF-8 - degrades sulphate
PT groups of proteoglycan sugar chains, useful for treating bone
PT metabolic diseases
PS Claim 2; Page 17-22; 29pp; English.

CC cDNA of mouse OSF-8 (mOSF-8) was isolated from mouse osteoblastic cell line MC3T3-E1 cDNA library by a combination of PCR and the subtraction method, and by the differential screening method. OSF-8 has homology with known representative sulphatase molecules but belongs to a new subclass different from those so far reported. The inventors claim the precursor protein (R51355 AAs 1-556) which includes the signal peptide (AAs 1-18), and they also claim a protein comprised of residues 19-556. They also claim DNA or RNA encoding the claimed proteins.

SQ Sequence 556 AA;
DB 9; Score 172; Match 37.2%; QryMatch 4.2%; Pred. No. 2.75e+05; Matches 29; Conservative 19; Mismatches 28; Indels 2; Gaps 2;

Db 274 knirafyamaedaamlgeillalhlkldllqktivtytdhg-emamehrfykmsmye 332

Qy 294 RKIRQSFASVSYLDTVGRLSALDQLANSTIIAFTSDHGWLG-EGEWAKYSNFD 352

Db 333 asvhvpllmqgqikanl 350

Qy 353 VATHVPLIFVPGRTASL 370

RESULT 2

ID R06341 standard; protein; 1671 AA.

AC R06341;

DT 14-DEC-1990 (first entry)

DE Signal peptide and first 1643 AAs of mature apoB protein.

KW Atherosclerosis; lipid binding peptide; LBP; phospholipid;

KW Lecithin cholesterol acyl transferase; L-CAT; apo A1;

KW Chinese hamster ovaries;

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..27

FT /label=Signal peptide.

FT /label=Mature peptide.

PN US4943527-A.

PD 24-JUL-1990.

PF 27-FEB-1986; 834300.

PR 04-OCT-1985; US-784418.

PR 04-DEC-1985; US-804692.

PR 27-FEB-1986; US-834300.

PA (CALB-) CALIFORNIA BIOTECHN.

PI Protter AA, Vigne JL, Mallory JB, Talmadge KD, Kane JP;

DR MPI; 90-246622/32.

DR N-PSDB; 005590.

PT Prodn. of purified lipid-binding peptide - capable of binding to

PT phospholipid(s), useful in e.g. atherosclerosis.

PS Disclosure; pp; English.

CC Lipid-binding peptides in conjunction with phospholipids are

CC effective in reversing cholesterol transport, and may enhance

CC lecithin cholesterol acetyl transferase, lowering the rate of

CC plaque formation and atherosclerosis.

SQ Sequence 1671 AA;

DB 1; Score 110; Match 27.3%; QryMatch 2.7%; Pred. No. 1.74e+00;

Matches 24; Conservative 22; Mismatches 37; Indels 5; Gaps 5;

Db 1292 nsaki-eipilpggksrdlkmletvrtpalhf-ksvghlpserefqvptftipklyqlq 1349

Qy 187 DVLDPVECTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPK-EFQKLYPLE 245

Db 1350 -vpilgvlstnsvslyn-wsgllqw 1375

Qy 246 NITLAPDEVPDGLPPVAYNPWMDIRQR 273

RESULT 3

ID P70647 standard; protein; 2721 AA.

AC P70647;

DT 17-APR-1991 (first entry)

DE Sequence of N-terminal apolipoprotein B (apoB).

KW Lipid-binding peptide; apolipoprotein purification.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 87..167

FT /label=Signal

FT Protein 168..8247

PN M08702062-A.

PD 09-APR-1987.

PF 02-OCT-1986; 002075.

PR 04-OCT-1985; US-784418.

PR 04-DEC-1985; US-804692.

PR 27-FEB-1986; US-834300.

PA (BIOT-) BIOTECHN RES PARTN.

PI Protter AA, Vigne JL, Mallory JB, Talmadge KD;

DR MPI; 87-108703/15.

DR N-PSDB; N70996.

PT Purified lipid-binding peptide prodn. - by expression of

PT recombinant system and adding lipid to obtd. peptide mixt. to

PT form lipo-peptide complex

PS Claim 50; pp73-85; 123pp; English.

CC ApoB is a lipid-binding peptide adapted to bind to phospholipids at

CC one or more amphipathic alpha-helical peptide regions. ApoB gene is

CC pref. expressed in Chinese hamster ovary (CHO) cells and the gene is

CC introduced in an expression vector having a regulatable promoter

CC derived from the human metallothionein II gene.

SQ Sequence 2721 AA;

DB 2; Score 110; Match 27.3%; QryMatch 2.7%; Pred. No. 1.74e+00;

Matches 24; Conservative 22; Mismatches 37; Indels 5; Gaps 5;

Db 1292 nsaki-eipilpggksrdlkmletvrtpalhf-ksvghlpserefqvptftipklyqlq 1349

Qy 187 DVLDPVECTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPK-EFQKLYPLE 245

Db 1350 -vpilgvlstnsvslyn-wsgllqw 1375

Qy 246 NITLAPDEVPDGLPPVAYNPWMDIRQR 273

RESULT 4

ID R44211 standard; protein; 351 AA.

AC R44211;

DT 26-MAY-1994 (first entry)

DE Caffeine demethylase.

KW Caffeine demethylase; CDM; Pseudomonas; 3-methyl-7-alkyl-xanthine;

KW 1,3-dimethyl-7-alkyl-xanthine; demethylation; ds.

OS Pseudomonas sp.

PN DE4316882-A.

PD 25-NOV-1993.

PF 19-MAY-1993; 316882.

PR 20-MAY-1992; JP-154380.

PR 27-OCT-1992; JP-312954.

PA (AMAN) AMANO PHARM KK.

PI Imai Y, Koide Y, Nakane S;

DR MPI; 93-378610/48.

Result	Query			ID	Description	Pred. No.
	No.	Score	Match Length			
1	4069	100.0	550	9	iduronate-2-sulfatas	0.00e+00
2	3129	76.9	563	10	iduronate-2-sulfatas	0.00e+00
3	1300	31.9	212	10	iduronate-2-sulfatas	2.47e-196
4	213	5.2	583	9	steryl-sulfatase (EC	4.55e-14
5	211	5.2	567	9	steryl-sulfatase (EC 3	8.76e-14
6	211	5.2	522	9	arylsulfatase (EC 3.1	8.76e-14
					N-acetylglucosamin	

KEYWORDS
SUMMARY
sulfuric ester hydrolase
#length 550 #molecular-weight 61872 #checksum 5247

[illegible]

```

RESULT      3
ENTRY       PN0565      #type fragment
TITLE       iduronate-2-sulfatase (EC 3.1.6.13) (clone MTA13) - mouse
            (fragment)
ALTERNATE_NAMES iduronate sulfatase
ORGANISM     #formal name Mus musculus #common name house mouse
DATE         05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change
            05-Aug-1994
ACCESSIONS  PN0565
REFERENCE    PN0565
AUTHORS      Daniele, A.; Russo, T.; Ballabio, A.; Di Natale, P.
#journal     Biochem. Biophys. Res. Commun. (1993) 194:1030-1037
#title       The mouse iduronate sulfatase gene: Identification of a novel
            transcript.
#accession   PN0565
#molecule_type mRNA
#residues    1-212 #label DAN

```

17 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050

ALTERNATE NAMES N-acetylglactosamine-6-sulfate sulfatase

Result	Query			DB	ID	Description	Pred. No.
	No.	Score	Match Length				
1	4069	100.0	550	4	IDS_HUMAN	IDURONATE 2-SULFATASE	0.00e+00
2	3129	76.9	563	4	IDS_MOUSE	IDURONATE 2-SULFATASE	0.00e+00
3	282	6.9	497	8	YIDJ_ECOLI	HYPOTHETICAL 57.3_KD	4.30e-32
4	213	5.2	583	7	STS_HUMAN	STERYL-SULFATASE PREC	6.94e-19
5	211	5.2	522	3	GAGS_HUMAN	N-ACETYLGALACTOSAMINE	1.62e-18
6	195	4.8	507	1	ARSL_HUMAN	ARYLSULFATASE A PRECU	1.29e-15
7	190	4.7	533	1	ARSB_HUMAN	ARYLSULFATASE B PRECU	1.01e-14
8	184	4.5	551	1	ARS_HEMPU	ARYLSULFATASE PRECURS	1.15e-13
9	172	4.2	535	1	ARSB_FELCA	ARYLSULFATASE B PRECU	1.39e-11
10	153	3.8	551	1	ASLA_ECOLI	ARYLSULFATASE (EC 3.1	2.10e-08

Db	301	faevsyldtqvrllaealddqlanetiafstdhgalgehwakysnfadvathvpli	360
Qy	301	fasvysldtqvrllsalddqlanetiafstdhgalgehwakysnfadvathvpli	360
Db	361	fyvgptaslpaeaklfpyldpfdasqmpgrgsmldvelvslfptlaglaqlvpp	420
Qy	361	fyvgptaslpaeaklfpyldpfdasqmpgrgsmldvelvslfptlaglaqlvpp	420
Db	421	rcpvpsfhvelcregknlkhfrdledpylpgnpreliaysqyrpsdipqwnskp	480
Qy	421	rcpvpsfhvelcregknlkhfrdledpylpgnpreliaysqyrpsdipqwnskp	480
Db	481	slkdikimgysirtidyrytwvgnpdeflanfsdihagelyfyvdsdplqdhmnyndsq	540
Qy	481	slkdikimgysirtidyrytwvgnpdeflanfsdihagelyfyvdsdplqdhmnyndsq	540
Db	541	ggdlfqllmp	550
Qy	541	ggdlfqllmp	550
RESULT 2			
ID	IDS	MOUSE	STANDARD; PRT; 563 AA.
AC	Q08890;		
DT	01-OCT-1994 (REL. 30, CREATED)		
DT	01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)		
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)		
GN	IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).		
DE	IDS.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; RODENTIA.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=THYMUS;		
RM	9315172		
RL	DANIELE A., FAUST C.J., HERMAN G.E., DI NATALE P., BALLABIO A.;		
RA	GENOMICS 16:755-757(1993).		
CC	-1- FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN		
CC	SULFATE AND DERMATAN SULFATE.		
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE		
CC	L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE		
CC	AND HEPARIN.		
CC	-1- SUBCELLULAR LOCATION: LYSOSOMAL.		
CC	-1- SIMILARITY: TO OTHER SULFATASES.		
DR	EMBL; L07921; MWIDS.		
DR	PIR; A47153; A47153.		
KW	HYDROLASE; GLYCOPROTEIN; LYSOSOME; ZYMAGEN; SIGNAL.		
FT	SIGNAL 1 ?		
FT	PROPEP ? 59		
FT	CHAIN 60 563		
FT	CARBOHYD 130 130		
FT	CARBOHYD 159 159		
FT	CARBOHYD 261 261		
FT	CARBOHYD 295 295		
FT	CARBOHYD 526 526		
FT	CARBOHYD 550 550		
SQ	SEQUENCE 563 AA; 63437 MW; 1700395 CN;		
DB 4; Score 3129; Match 79.0%; QryMatch 76.9%; Pred. No. 0.00e+00			
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps			
Db	19	arairwqlsfellgscfaleaagcnatdnlillivddlrpslcygcklvrsnp	78
	:	:	:

RESULT	3
ID	YIDJ ECOLI STANDARD; PRT; 497 AA.
AC	P31447;
AD	01-JUL-1993 (REL. 26, CREATED)
DT	01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DDT	01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE	HYPOTHETICAL 57.3 KD PROTEIN IN ILVO-IBPB INTERGENIC REGION.
GN	YIDJ.
OS	ESCHERICHIA COLI.
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC	ENTEROBACTERIACEAE.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MG1655;
RM	93315143
RA	BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RL	GENOMICS 16:551-561(1993).
CC	-/- SIMILARITY: STRONG, TO SULFATASES.
DR	EMBL; L10328; ECOW82.
DR	ECOGENE; EG11705; YIDJ.
DR	PROSITE; PS00149; SULFATASE_2.
DR	PROSITE; PS00523; SULFATASE_1.
KW	HYPOTHETICAL PROTEIN; HYDROLASE.
FT	ACT SITE 102 102 POTENTIAL.
SQ	SEQUENCE 497 AA; 57295 MW; 1238485 CN;

RESULT	4	STANDARD;	PRT; 583 AA.
ID STS HUMAN			
AC	P08842;		
DT	01-NOV-1988 (REL. 09, CREATED)		
DT	01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)		
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)		
DE	STERYL-SULFATASE PRECURSOR (EC 3.1.6.2) (STEROID SULFATASE) (STERYL-SULFATE SULFATASE)		
DE	SULFATE SULFATASE (EC 3.1.6.2) (STEROID SULFATASE) (STERYL-SULFATE SULFATASE)		

ID	STS HUMAN	STANDARD;	PRT; 583 AA.
DT	P08842;		
DT	01-NOV-1988 (REL. 09, CREATED)		
DT	01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)		
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)		
DE	STERYL-SULFATASE PRECURSOR (EC 3.1.6.2) (STEROID SULFATASE) (STERYL-SULFATE SULFOHYDROLASE) (ARYLSULFATASE C).		
DE	STS.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; PRIMATES.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RM	89340479		
RA	STEIN C., HILLE A., SEIDEL J., RUNBOUT S., WAHEED A., SCHMIDT B.,		
RA	GEUZE H., VON FIGURA K.;		
RL	J. BIOL. CHEM. 264:13865-13872(1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RM	87187642		
RA	YEN P.H., ALLEN E., MARSH B., MOHANDAS T., WANG N., TAGGART R.T.,		
RA	SHAPIRO L.J.;		
RL	CELL 49:443-454(1987).		
RN	[3]		
RP	SEQUENCE OF 22-45.		
RP	TISSUE=LIVER;		
RM	89352671		
RA	KAWANO J.-I., KOTANI T., OHTAKI S., MINAMINO N., MATSUDO H., OINUMA T.,		
RA	AIKAWA E.;		
RL	BIOCHIM. BIOPHYS. ACTA 997:199-205(1989).		
RN	[4]		
RP	VARIANTS XLI LEU-341; ARG-372 AND TYR-446.		
RM	92170784		
RA	BASLER E., GROMPE M., PARENTI G., VATES J., BALLABIO A.		